From: *Rational Models of Cognition*, M. Oaksford and N. Chater (eds). Pages 441-468, Oxford University Press, Oxford, UK, 1998.

The Rational Analysis of Inquiry: The Case of Parsing

Nick Chater*
Department of Psychology,
University of Warwick.

Matthew Crocker
Centre for Cognitive Science
University of Edinburgh

Martin Pickering
Department of Psychology
University of Glasgow

Word Count: 9000 words

E-mail : nick.chater@warwick.ac.uk
martin@psy.gla.ac.uk
mwc@cogsci.ed.ac.uk

Tel : +44 01203 523537

*Please address correspondence concerning this article to Nick Chater, Department of Psychology, University of Warwick, Coventry CV4 7AL, UK, Martin Pickering, Department of Psychology, University of Glasgow, 56, Hillhead Street, Glasgow, G12 9YR, UK, or Matthew Crocker, Centre for Cognitive Science, 2, Buccleuch Place, Edinburgh, EH8 9LW, UK.

Acknowledgement. We would like to thank Mike Oaksford for valuable comments on a previous version of this paper.

The cognitive system is not merely a passive receiver of information. It has some measure of control of what information it receives; and how that information is processed. Control over the information received may be exercised in a wide variety of ways, from adjustments to the sense organs (e.g., by moving the eyes), to decisions concerning which newspaper to read. Control over how information is processed is equally ubiquitous, ranging from attentional mechanisms (presuming that such mechanisms at least to some degree bias the resources applied to processing different aspects of the sensory input) to how much effort to spend thinking about a move in a possible move in a chess game, or on a decision in everyday life.

The cognitive system is engaged in a process of inquiry[1] about the world: it must chose how to collect information and how to process that information. An analogy is intended here between inquiry by the cognitive system and organized inquiry involved in research. Inquiry in research also involves judicious control over information (e.g., which experiments are performed, which observations are made) and how that information is processed (how that data is analysed, what hypotheses are proposed, when are they abandoned, and so on). We shall develop this analogy further below.

Inquiry is difficult, whether in a cognitive or scientific context, because it typically must proceed in the face of very severe resource limitations. Information <u>gathering</u> must therefore be highly selective. It is not possible, for example, to obtain a high-resolution image of the entire visual world; the fovea must be directed towards a tiny part of that image at any one time. Similarly, it is not possible to read all newspapers, or all articles in a single newspaper; reading must be highly selective to obtain maximum benefit. Information <u>processing</u> must also be highly selective: attention cannot be directed everywhere; only a tiny fraction of sequences of chess moves can be analysed, and so on.

Given that people (and animals) appear to be highly successful information processing systems, presumably these problem of directing inquiry are typically solved satisfactorily. How is this possible? This chapter provides a framework for answering questions of this form, at the level of <u>rational analysis</u> (Anderson, 1990, 1991; Oaksford & Chater, 1995a). Rational analysis involves providing a <u>normatively justified</u> account of how the problem can be solved, and this account must also be <u>descriptively correct</u> in that the cognitive processes under study must implement (or, more likely, cheaply approximate) this

[1] The use of “inquiry” in this sense was suggested to us by Mike Oaksford, in relation to work in his PhD thesis (Oaksford, 1989), who has also contributed significantly to many of the fundamental ideas in this paper. This term was used in a related sense by Stalnaker (1984).

analysis. The criterion of descriptive correctness ties the account to empirical data; the criterion of normative justification explains why the cognitive process is successful.

This chapter is concerned exclusively with outlining normative proposals for the rational analysis of inquiry. The main goal is to explore the range of options available in building potential rational analysis of specific aspects of cognition involving inquiry. We do this in two ways: first by outlining the theoretical underpinnings of Bayesian and decision-theoretic approaches to the rational direction of inquiry; and second by showing in detail how a rational analysis can be developed, taking the case of local ambiguity resolution in parsing. The descriptive correctness of potential accounts of this form, when related to empirical evidence, is outside the scope of this chapter. By focussing on normative issues, this paper has the flavour of a priori inquiry; but notice that the ultimate usefulness of any of the approaches to the rational analysis of cognitive processes involving inquiry depends crucially on the ability to explain empirical data, and thus satisfy both the normative and descriptive criteria.

This paper is divided into two parts, dealing with general issues and the specific rational analysis of parsing in turn. In Part 1, we begin by considering the general frameworks for rationality in terms of which normative accounts of inquiry can be developed. We then consider how these frameworks allow us to define the value of information. This allows us to consider how to estimate the value of information-gathering inquiry, which has the goal of obtaining information. Roughly, this is the expected value of the information that will be obtained, balanced against the costs of conducting the inquiry. Part 1 concludes with a discussion of the relationship between the different frameworks that have been introduced. In Part 2, we show how a specific rational analysis of inquiry can be developed, drawing on the problem of local ambiguity resolution in parsing. Because natural language is locally highly ambiguous, there are frequently a number of possible parses consistent with the current input; computational resource limitations presumably mean that all possibilities cannot be considered simultaneously. So the parser must choose in which order to consider the various possible parses consistent with the current input. We provide a rational analysis of how these choices should be made, assuming that the goal of the parser is to maximize the probability of obtaining the globally correct parse for the sentence. Finally, we briefly draw some conclusions concerning future research on the rational analysis of cognitive processes involving inquiry.

PART 1: A FRAMEWORK FOR THE RATIONAL ANALYSIS OF INQUIRY

Formal frameworks for rationality

Before we can develop a normative account of how inquiry should proceed, we need to establish a formal framework for rationality, in terms of which the rationality of decisions about inquiry can be assessed.

We consider two possible starting points. The first begins by searching for a framework for rational thought – the goal is to know as much as possible about the world. The second begins with the goal of rational choice--the goal is to make the best possible decisions. In the first framework, inquiry is valuable for its own sake, because it leads to knowledge--inquiry is "disinterested". In the second framework, inquiry is only worthwhile if it leads to better decisions.

Both starting points may seem unpromising because they appear to raise extremely deep issues. The first starting point, of pursuing the goal of knowledge, corresponds to facing the problem of epistemology; the second, of deciding what to do, corresponds to the problem of ethics.

Fortunately, however, both classes of problem can be reformulated in more modest and more tractable terms. The problem of pursuing knowledge can be reformulated as the problem of how to move from a given set of beliefs to further beliefs. The deepest questions of epistemology--concerning how beliefs can ultimately be grounded, rather than circularly depending on each other--are thus side-stepped. A paradigm approach to this reformulated problem is as follows. We assume that each proposition is associated with a number between 0 and 1, expressing the "degree of belief" in that proposition (where 0 denotes certainty that the proposition is false, and 1 denotes certainty that it is true). The problem of deciding what to think can now be viewed as a problem of deciding what number to associate to a given proposition, in the light of the numbers that one has already assigned to other propositions. There are a large number of well-known lines of arguments (see Earman, 1992; Howson & Urbach, 1989) which converge on the conclusion that this process should follow the standard laws of probability theory. These arguments legitimise the possibility of interpreting probability theory as a normative calculus for uncertain inference[2]. Specifically, probability theory can serve as the prescription for what we should

[2] Of course, there are also other legitimate interpretations of probability theory, such as expressing relations between limiting frequencies of events of various kinds in sequences of repeated experiments. Debates concerning the correct interpretation of probability typically centre on which interpretation is appropriate in scientific or statistical inquiry, but are not

think, given some (typically very partial) set of given information: this is commonly known as the Bayesian approach to inference (de Finetti, 1972; Keynes, 1921; Lindley, 1971).

The problem of making decisions can be reformulated as the problem of how to move from a given set of beliefs and desires to actions. The deepest questions of ethics--concerning how desires should ultimately be grounded, rather than circularly depending on each other--are thus also side-stepped[3]. A paradigm approach to this reformulated problem is as follows. We assume that a real number can be associated with each possible event (or "outcome"), representing its "subjective utility" for the decision maker[4]. Higher valued events are assumed to be preferred in a forced choice with lower valued events. Beliefs are associated with probabilities as before. There are then a number of well-known arguments which converge on the conclusion that choosing how to act should follow the laws of decision theory. The recommendation of decision theory is that actions should be chosen to maximize expected utility, where the measure of utility, and the probabilities in the light of which the expectation is assessed, are take as given[5].

These frameworks for rational choice are valuable starting points for the rational analysis of cognitive processes. But before launching into discussing how these frameworks may be applied, a few clarificatory comments are in order. First, recall that a rational analysis must be normatively justified and descriptively correct (Oaksford & Chaterr, 1995a, 1996). These rational choice models have normative justifications, but it is an empirical issue whether cognitive processes can fruitfully be viewed as providing approximations to them. This question can only be decided case-by-case by constructing rational analyses, and comparing them against relevant empirical data. As with any other style of explanation in science, a rational analysis will typically be iteratively refined in the light of empirical results; rather than being derived a priori. Second, note that we have framed the discussion so far in terms of agents, with beliefs, desires and actions. But the same style of analysis may be applied to sub-personal cognitive processes, as well as agents--indeed this is the standard case for the rational analysis of cognition. In such cases, "beliefs" are information states (associated with numerical values interpretable as

---

relevant in the psychological context of developing rational analyses, where the subjective interpretation is the only meaningful possibility.

[3] We also ignore the case in which the rational agent must reason about the behaviour of other rational agents, which is treated by game-theory and discussed in Colman, this volume.

[4] This requires that a subjective utility function exists for the person's set of preferences (e.g., de Groot, 1970).

[5] Though it may also take other forms, such as the "minimax" criterion, which recommends choosing the action whose worst case is least bad, which is standard in game theory (von Neumann & Morgenstern, 1944).

probabilities), "desires" or "utilities" are simply defined by (typically) numerical values; actions correspond to operations or outputs of the processor. The rational analysis of parsing developed will have this character. For example, it assumes that the parser chooses operations to maximize the probability of finding the intended global parse of the sentence. But this does not mean that the parser really has beliefs or desires about parses, or whether they succeed, merely that the parser is assumed, for example, to store numbers which represent the prior probability it assigns to various parsing structures. Indeed, even this assumption is dispensable, because the prescriptions of the rational analysis may be implemented (or approximated) using all kinds of different algorithms and representations. Nonetheless, for expository clarity, we will talk of cognitive processes "deciding," "choosing," "preferring," and so on--such usages are dispensable, although convenient, anthropomorphisms.

This brings us to the third point of clarification: that the use of rational theories such probability theory and decision theory in the rational analysis of some of the cognitive <u>mechanisms</u> of which agents are composed in no way implies that the <u>whole agent</u>'s beliefs, desires and actions can be given a rational analysis. The validity of such "rational choice" explanations of human behaviour as a whole is, of course, a question of fundamental importance, which has attracted extensive controversy within experimental psychology (for example, Gigerenzer, Hoffrage & Kleinbölting, 1991; Kahneman, D, 1996; Kahneman, Slovic & Tversky, 1982), as well as in economics and the social sciences (see, for example, Arrow, Colombatto, Perlman & Schmidt, 1996; Elster, 1983, 1989; Sen, 1990; Simon, 1982). But the usefulness of the rational analysis of cognitive mechanisms is independent of such debates.

We noted at the outset that choices about what to do or think must typically be made in the face of limited information, and limited computational resources[6]. The relative emphasis on these two kinds of limitation will, of course, depend on the context. In some contexts, the computational resources may be sufficient to process the available, partial, data, and the main question of interest is how to choose which new data to select. Here, the problem is <u>optimal data selection</u>. In other contexts, the main problem may be limitations on computational resources available to use existing data, and the main question of interest is how to allocate these resources optimally to analyse this data. Here the problem is <u>optimal computational resource allocation</u>. There are, of course, hybrids where both factors are important, but we shall ignore these for simplicity.

[6] This corresponds to the general distinction between data-limited and resource-limited cognitive processes (Norman & Bobrow, 1975).

In this paper, we present a uniform treatment of problems. The first problem, optimal data selection, has been extensively studied; but the second, optimal computational resource allocation, is relatively unexplored. But the two problems are closely related--in essence the decision to carry out some further calculation can be viewed as directly analogous to the decision to collect some more data--and hence the analyses of the first problem can shed light on the second. Carrying out observations and carrying out calculations are simply two different types of inquiry which can lead to new information (although, as we shall note below, the problem of computational resource allocation does raise certain fresh issues).

We now consider how the two frameworks for rational choice introduced above can be used to provide theories of how inquiry should proceed. This requires first establishing how each framework allows a value to be attached to information.

The value of information

From the perspective of "disinterested inquiry," a natural and straightforward goal is simply to maximize the amount of information gained[7] by an investigation, whether or not this data is useful with respect to other utilities (Lindley, 1956; Mackay, 1992). Specifically, the amount of information gained is viewed as reduction in uncertainty, where uncertainty is measured by the standard entropy measure from information theory (Shannon, 1948). We shall consider this measure in the analysis below, partly because of its generality, and partly because it has been analysed mathematically. There are, however, many more specific informational goals which may be more relevant to the rational analysis of particular cognitive tasks, particular where more general goals do not lead directly to a tractable analysis. We shall choose one such specific goal in developing a rational analysis of local ambiguity resolution in parsing below – namely, maximizing the probability of correctly parsing a sentence.

---

[7] An alternative goal to maximizing information gain is maximizing the change from the distribution of probabilities over the hypotheses before and after the data is received. The standard measure of this change is the Kullback-Liebler distance between the distribution in the light of the data and the original distribution: $\sum_i P(H_i|K\&N)\log_2\left(\frac{P(H_i|K\&N)}{P(H_i|K)}\right)$, where H_i, K and N have the meanings outlined below in the text. It turns out that the expected value of this quantity is identical to the expected value of information gain, even though information gain and Kullback-Liebler distance are not the same for any particular piece of data, N (Lindley, 1956). Thus, we have an alternative starting point for an account of optimal data selection. This may have advantages in certain contexts (see Oaksford and Chater's (1996) response to Evans and Over (1996) for an example in the context of a rational analysis of the selection task).

We assume that the agent[8] starts with background knowledge, K. This knowledge determines the degree of belief in a range of mutually exclusive and exhaustive hypotheses, H_i, from a set, H, such that each H_i is associated with a probability, given the initial background knowledge, K: $P(H_i|K)$. The agent's aim is to learn more about which of these hypotheses is true. What is the value associated with a particular piece of new information, N (which may be the result of some investigation, or might be given to the agent "for free")?

On this approach, the value of N is determined by the amount of information that the agent gains about H. Information gain is the initial uncertainty minus the revised uncertainty, i.e., after N is known:

The agent's initial uncertainty, Uncertainty(H|K) is:

$$\text{Uncertainty}(H|K) = -\sum_i P(H_i|K)\log_2 P(H_i|K) \quad (1)$$

After N is known, the probabilities of the H_i will be revised, using Bayes' theorem[9] to $P(H_i|K\&N)$, and the uncertainty will then be:

$$\text{Uncertainty}(H|K\&N) = -\sum_i P(H_i|K\&N)\log_2 P(H_i|K\&N) \quad (2)$$

The information gain, $I_g(N)$, associated with N is therefore expressed:

$$I_g(N) = \text{Uncertainty}(H|K) - \text{Uncertainty}(H|K\&N) \quad (3)$$

This is the value of the information, N, with respect to the set of hypotheses, H, from the point of view of disinterested inquiry.

We now turn to the value of information from the point of view of decision making (see Berger, 1985). The agent must choose between a set of actions, A_i. These actions are assumed to have an impact on some aspect of the world, which is of interest to the agent. This aspect of the world is modeled as a discrete set of outcomes, Out_j, which is associated with a utility $U(Out_j)$. For example, the actions might correspond to the choice of the kind of flour used when baking a cake, and the outcome might be the quality of the cake (good,

---

[8] The use of the term "agent," like "knowledge," is merely for convenience, to refer to a reasoning system of some kind. As we mentioned above, there is no commitment to whatever aspect of the cognitive system is under study having beliefs, desires, and the like.

[9] Specifically, $P(H_i|K\&N) = \dfrac{P(N|H_i\&K)P(H_i|K)}{\sum_j P(N|H_j\&K)P(H_j|K)}$.

average, or poor), according to some criterion. But the agent may not know what the relation between the choices and outcomes is. The agent's partial information about the relationship between actions and outcomes may be expressed by the agent's estimates of the conditional probabilities of each kind of outcome, given each kind of action: $P(\underline{Out_j}|\underline{A_i}, \underline{K})$. Notice that the presence of the "$\underline{K}$" in this formula captures the fact that these estimates are determined in the light of background knowledge, excluding the new information.

These estimates allow the expected utility, $\underline{EU}(\underline{A_i}|\underline{K})$, of each action to be calculated, given knowledge $\underline{K}$. Specifically,

$$EU(A_i|K) = \sum_j P(Out_j|A_i, K)U(Out_j) \tag{4}$$

The best policy is to choose the action, $\underline{A_m}$, which maximizes this quantity, i.e., where

$$m = \arg\max_i EU(A_i|K) \tag{5}$$

and thus,

$$EU(A_m|K) = \max_i \left(\sum_j P(Out_j|A_i, K)U(Out_j) \right) \tag{6}$$

Now, after the new information $\underline{N}$ arrives, the agent can make a new set of estimates of the relationship between actions and outcomes, taking account of this new information: $P(\underline{Out_j}|\underline{A_m}, \underline{K}\&\underline{N})$. This allows the agent to produce a revised estimate of how successful the <u>previous</u> choice of strategy would be, using the action $\underline{A_m}$, chosen by the strategy outlined above.

$$EU(A_m|K\&N) = \sum_j P(Out_j|A_m, K\&N)U(Out_j) \tag{7}$$

So this gives an estimate (in the light of now knowing $\underline{N}$) of how well the agent would have done before $\underline{N}$ was known. To obtain an estimate of the <u>value</u> of $\underline{N}$, we need to contrast this figure with the estimate of the expected utility if the action is chosen <u>after</u> information $\underline{N}$ has arrived. Paralleling (7), we note that the expected utility of arbitrary action, $\underline{A_i}$, is now estimated as:

$$EU(A_i|K\&N) = \sum_j P(Out_j|A_i, K\&N)U(Out_j) \tag{8}$$

The best policy is to choose the action, A_q, which maximizes this quantity, i.e., where

$$q = \arg\max_i EU(A_i|K\&N) \tag{9}$$

and thus,

$$EU(A_q|K\&N) = \max_i \left(\sum_j P(Out_j|A_i, K\&N)U(Out_j) \right) \tag{10}$$

Thus, the gain in expected utility, i.e., the value, $V(N)$, associated with learning N is:

$$V(N) = EU(A_q|K\&N) - EU(A_m|K\&N) \tag{11}$$

where q is the choice of action in the light of N, and m is the choice of action before m is known. Note, of course, that if N does not cause the agent to change the choice of action, the two terms on the right hand side of (8) are the same, and $V(N) = 0$. So, in sum, the value of a piece of new information is determined by the impact that it has on the agent's expectations about what will occur given each action; but this impact is only of value if the agent actually decides to change course of action. This follows the spirit of the utility-based approach--information is not of interest for its own sake, but only in its influence on action[10].

In this section, we have outlined measures of the value of information, one from the perspective of disinterested inquiry, and one where information is to be used to maximize expected utilities. We now consider how these accounts of the value of information can be used as the basis for calculating the expected value of the information that will be obtained from a process of inquiry.

The expected value of inquiry

[10] As stated, the utility-based view of information may seem unacceptably myopic. After all, it is possible that information may not immediately cause a change of action, but when combined with later information, it may do so. The utility-based view can be extended to take account of sequential information gathering, essentially by taking expectations about the possible results of future inquiry, their impact on action, and hence on utility. This leads to the field of "preposterior" analysis (e.g., Berger 1985) where calculations are extremely complex both conceptually and in terms of computational tractability.

Now we have defined the value of a piece of information, we can now consider the expected value of inquiry[11]. We begin by putting aside a worry: that the very idea of choosing a line of inquiry to give the most valuable information is conceptually incoherent, whatever the criterion of value. Only if one has obtained some data can one know how it is relevant to the problem in hand; and therefore it seems impossible to predict beforehand which data is likely to be useful. Fortunately, this argument is not valid. Although it is not possible to predict how useful the outcome of a particular search for more data will be, it is possible to estimate the expected value of this information, based on one's current knowledge. Optimal inquiry involves choosing to obtain new data so that the expected value of the information obtained is as great as possible.

The theory of optimal data selection develops in two different ways, depending on which of the two notions of the value of information described above is used. If value is considered as a utility of some kind, which is to be traded off against other utilities, then we can set the problem in a decision-theoretic context. This gives rise to the extensive theory of so-called "preposterior analysis" in Bayesian decision theory (e.g., Berger, 1985; Wald, 1947, 1950). If the value of information is not considered as a utility, but value is measured instead as the amount of information gained, then the problem may be set in terms of probability theory and information theory. This gives rise to the Bayesian analysis of the information value in an experiment (Lindley, 1956; Good, 1966; Mackay, 1992).

We now consider the formal development of the disinterested and decision-theoretic approaches to assessing the expected value of inquiry in turn.

From a disinterested perspective, inquiry should be directed to have the greatest possible expected information gain. Suppose that we are considering an investigation, $\underline{Inv}$, which itself has cost $\underline{C}(\underline{Inv})$[12]. Whether it is worth carrying out the investigation $\underline{Inv}$ depends on whether the expected value of the information gathered by $\underline{Inv}$ exceeds the cost of the investigation $\underline{C}(\underline{Inv})$.

The value of the information that will be obtained by $\underline{Inv}$ cannot be known before the investigation. This is because the result, $\underline{Inv_{result}}$, cannot be known until the investigation has been carried out. However, the agent will have expectations about the results, $P(\underline{Inv_{result}}|K)$. It is therefore possible to calculate the expected value of the investigation, based on these expectations, before it is carried out:

The expected information gain of $\underline{Inv}$, $\underline{EI_g}(\underline{Inv})$, is the sum of the information gains associated with each possible result, weighted by the probability of that result:

[11] See also Young (this volume) for a related analysis of the value of inquiry in the context of understanding how people search computer menus.
[12] In general, the cost may not be fixed, but may also need to be estimated.

$$EI_g(Inv) = \sum_{result} P(Inv_{result}|K) I_g(Inv_{result}) \qquad (12)$$

Given a choice between different mutually exclusive courses of investigation, the agent should choose the investigation Inv with the highest EI_g score.

This measure of the value of disinterested inquiry has been recently used as the basis for a rational analysis of so-called "indicative" versions of Wason's (1966, 1968) selection task (Oaksford & Chater, 1994, 1995b, 1996; see also Oaksford & Chater, this volume; Over & Jessop, this volume). In "indicative" selection tasks, people are asked to test whether a hypothesis is true or false. This suggests that people may understand this task as involving disinterested inquiry, because no utilities are either specified or suggested. In the context of the selection task, the possible "inquiries" are the turning over of each of four cards, and the hypotheses under consideration are a conditional rule and an "independence rule" about what is on each side of the cards. This rational analysis provides a good fit with a wide range of empirical data, and novel predictions of this account have recently been confirmed (Oaksford, Chater, Grainger & Larkin, in press).

Notice that this approach has entirely ignored the cost of carrying out an investigation--because the goal is disinterested inquiry, there is no direct connection between the value of the information obtained and the costs required to obtain that information. This is because information gain is measured in "bits" of information, whereas cost is measured in terms of utility of some kind, and hence these cannot meaningfully be compared. Nonetheless, costs can be taken into account as constraints on the investigation involved in disinterested inquiry, assuming that a finite amount of resources, R, is to be devoted to the pure pursuit of knowledge. Suppose that we have a number of possible lines of investigation, Inv_i, with costs $C(Inv_i)$, and with resources R. If we assume that the expected information gain from each Inv_i are independent, then the optimal portfolio of Inv_i is that which maximizes the overall information gain, while the sum of the $C(Inv_i)$ is less than R (i.e., while staying "within budget"). But notice that, in general, the information gain associated with different Inv_i will not be independent, because the results of one investigation will typically reduce the uncertainty about the results of other investigations. The mathematics required to deal with this is very complex, and the calculations will not typically be tractable.

We now consider the analysis for the case where the goal of inquiry is to improve decision making. Again, because the value of the information from the inquiry that will be obtained by <u>Inv</u> cannot be known before the investigation, the choice of inquiry must be based on the <u>expected</u> value of the investigation <u>EV</u>(<u>Inv</u>):

$$EV(Inv) = \left(\sum_{result} P(Inv_{result} | K) V(Inv_{result}) \right) - C(Inv) \tag{13}$$

Notice that, because the value of inquiry is measured in terms of utility, and the cost of the inquiry can also be measured in terms of utility, the cost can be incorporated directly into the equation, in contrast to the case of disinterested inquiry.

If the agent must choose between a range of possible investigations (typically including the "null" investigation, which gains no information, but costs nothing), the investigation, <u>Inv</u>, with the highest value <u>EV</u>(<u>Inv</u>) should be chosen.

This type of approach has been used by Anderson (1990) in proposing a rational analysis of problem solving – where inquiry consists of a search through a problem space. Anderson sees the key question in problem solving as searching this space to gain maximum advantage in the face of heavy computational costs. Oaksford and Chater (1994) also apply this type of analysis to explaining deontic selection tasks: i.e., selection tasks in which the participant must determine whether or not outcomes or behavior conform to some <u>normative</u> rule (e.g., that only people over 18 are allowed to drink). Normative rules cannot, of course, be tested. Instead, following Manktelow and Over (1987), Oaksford and Chater assume that people implicitly impose utilities concerning the importance of uncovering or missing violations of the rule (under-age drinkers). The expected utility of each card selection accurately predicts performance over a wide range of variants of the deontic selection task. Alongside this psychological work, there is also a well-developed literature on optimal investigation in zoology, where maximizing utilities such as food supply or frequency of mating is the goal (e.g., Kamil, Krebs & Pulliam, 1987; McNamara & Houston, 1992). The success of these rational analyses suggest that expected utility might provide the basis for successful rational analyses in other psychological domains.

The analyses we have given apply to both kinds of inquiry: information gathering and information processing. But notice that in the case of information processing with limited computational resources there is an additional complication: Deciding which information to process may <u>itself</u> take up significant computational resources. Therefore some balance must be struck between, on the one hand, spending so much computational effort deciding how the limited computational resources should be used that the resources themselves are completely exhausted once the decision is reached, and on the other hand, leaping blindly into some particular piece of information processing, with no notion whether this is likely to be fruitful.

A different, though also complex, case, arises where resources are not strictly limited, but must be obtained at some cost. For example, the time a person can spend deciding what to do may be unlimited in some contexts. The person could think all day and the next day and so on without reaching a decision--there is no rigid time limit. But by spending this time thinking, the person is not able to engage in some other enjoyable or useful activity. In economic terms, the time used for thinking must "bought" at the "cost" of the lost opportunity to do something else. As with the case where resources are rigidly limited, some balance must be found between being lost forever in thought, and choosing what to do completely blindly.

We do not attempt here to survey the possible ways in which these trade-offs may be made. The number of theoretical options is very large and unexplored, and which approach is adopted will depend on the cognitive process under consideration. In developing our rational analysis of local ambiguity resolution below, we shall describe a way of simplifying the problem of dealing with computational resource limitations: specifically, by assuming that the more resources are used, the more likely the computation is to "crash," and then trading off the expected probability of solving the problem by performing some computation against the probability that the computation will cause a "crash." An obvious way to set up such a trade-off is to have the goal of maximizing the overall probability of solving the problem successfully. We shall see how this approach can be applied in detail below. Note that this simplification is a general strategy in rational analysis. It will often be necessary, as here, to move from a very general goal where analysis is not tractable (e.g., maximizing expected information gain or expected utility given limited computational resources), to a specific goal where analysis is tractable, which is assumed to approximate the general goal. The general analysis is, nonetheless, important, because it defines a standard against which the appropriateness of the specific goal can be assessed.

Relation between normative accounts of inquiry for action and thought

We have sketched two approaches to optimal data selection, depending on whether the goal is optimizing action or optimal disinterested inquiry. It might seem that the second approach is redundant, at least from a psychological point of view, because "distinterested thought" might appear to be a rather implausible goal for a cognitive system. This line of thinking would suggest that the right goal for the cognitive system is some form of utility. From an evolutionary perspective, one might propose inclusive fitness; from an ethological perspective, one might propose maximizing goals such as nutritional value of diet, number of offspring, etc; from an economic perspective, one might propose maximizing some subjective notion of utility (as defined by the participant's preferences). Knowledge, it might seem, is only valuable to the extent that it contributes to guiding the choice of actions so that such goals as these are maximized.

In practice, however, an exclusive focus on utility may sometimes be inappropriate. Frequently, it is not feasible for a cognitive process or agent to decide what practical goal might be served by the results of some investigation. In science, for example, it is well-accepted that pure research may have enormous long-term practical implications, and is therefore worth funding; but in pure research the immediate task is disinterested inquiry--in our terms, attempting to maximize information gain, with no concern for potential uses to which that information might be put. From a utility-based perspective, the ideal would, of course, be to fund only that pure research with long term benefits. But it is not possible to decide beforehand which research this will be--and hence funding <u>dis</u>interested inquiry is an appropriate surrogate goal. Similarly, many cognitive processes may also be guided by the goal of "disinterested inquiry." Exploratory behavior in animals and children, and the general phenomenon of curiosity seem to suggest that this may be so. At the level of providing rational anlayses for cognitive mechanisms it is likewise possible that the appropriate goal is disinterested inquiry.

In general, a decision-theoretic perspective is likely to be appropriate in cases where the relation between information and action is relatively straightforward and relatively inflexible. Where this relation is complex and changeable, an information-gain rational analysis may be more appropriate.

Let us consider two examples from perception. First, consider the perceptual system of the frog, which is geared towards the detection of dark, fast, moving concave blobs (among other things) (Lettvin, Maturana , McCullough & Pitts, 1959). Thus, the frog's perceptual system seems adaptive not because it attempts to gain as much information about the world as possible, but because the information that it gains is relevant to its actions (e.g., snapping in the direction of the perceived blob), which relate to its utilities (e.g., eating flies). Notice that a decision-theoretic analysis seems appropriate here because the relation between the information concerning the motion of the blob is relatively straightforwardly and inflexibly related to something that has positive utility for the frog--eating a fly.

By contrast, the processes of human perceptual organization involved in making sense of line drawings, and other degraded stimuli commonly used in psychological studies, do not appear to relate directly or in a stable way to specific actions or utilities. Gestalt principles such as "good continuation" or "common fate" may aid the segmentation of the image, and assist in the process of object recognition; but the consequences that they have for action, if any, will depend critically on what it is that is being recognized. One and the same Gestalt principle may allow a perceiver in one case to recognize a lion, and in another to recognize a friend--the actions and utilities in the two cases will, of course, be radically different, and unpredictable. A myriad of background circumstances will also be relevant – in encountering a lion at the zoo, processes of perceptual organization contributing to recognition might have only the slightest utility and effect on actions, if the lion is behind bars; but might trigger flight and save one's life if the lion has escaped. The flexibility of the relationship between the information gained from processes of perceptual organization and their effect on utilities suggests that such processes might usefully be considered as primarily geared towards providing as rich a representation of the environment as possible: that is, as engaged in disinterested inquiry. Interestingly, the literature on perceptual organization has tacitly assumed that this is the goal of perceptual organization, by assuming that the goal of perceptual organization is utility-independent: that the goal is either constructing the most likely organization of the perceptual stimulus or the simplest (Chater (1996) has recently shown that, under natural interpretations of each of these viewpoints, they can be shown to be equivalent). The distinction between decision-theoretic and information-theoretic approaches provides an interesting framework for the general debate over the extent to which perception can be understood as having the goal of providing a rich general representation of the world (e.g., Marr, 1982), or as being geared towards serving particular actions (e.g., Gibson, 1979).

In the light of these considerations, it seems most appropriate to view the problem of parsing from the point of view of disinterested inquiry, rather than from a decision-theoretic perspective, given that the relation between information about the correct parse of a sentence may have an arbitrary, and typically extremely complex, relationship to the utilities and potential actions of the language understander. But a decision-theoretic analysis would seem to suggest that the parser should determine its strategies on the basis of exactly these factors. But while perhaps desirable in principle, such an analysis is intractably complex in practice. We therefore now develop a specific rational analysis of how the parser may choose to resolve syntactic ambiguities, based on an approximation to an information gain approach, rather than a utility-based approach.

PART 2: A CASE STUDY: LOCAL AMBIGUITY RESOLUTION IN PARSING

The problem of local ambiguity

We begin our rational analysis by outlining the problem that we aim to analyse: resolving local ambiguities in natural language parsing. In general terms, the sentence parsing task involves recovering an interpretation from a linguistic signal which is often highly ambiguous. A sentence may be globally ambiguous, as demonstrated by the following pair:

(a) “[NP: Flying planes] frightened the pilot”
(b) “[Gerund: Flying planes] frightened the child”

It is also possible, indeed very common, that sentences will contain local ambiguities. This arises from the fact that people process language incrementally, constructing a partial interpretation as each word is encountered, whether read or heard. This is demonstrated by the well-known reduced relative clause construction:

(c) “The actress sent the flowers was pleased”

When the word sent is encountered, the human parser may process it either as the main predicate of the sentence, and interpret the actress as its subject, or construct a reduced relative clause (Cf. "The actress who was sent...”). This is a local ambiguity, since ultimately only one or the other analysis will turn out to be correct. In this case, people systematically pursue the former analysis, and then garden-path when this ultimately turns out to be incorrect. If they cannot recover from this garden path, with some acceptable period of time, then we say that the parser has crashed. One mechanism for dealing with such ambiguity might be to construct all possible analyses in parallel. However, people’s inability to recover (often) from sentences such as (c), suggests that this is not the case. Indeed if we consider a more difficult example, we encounter further evidence against complete parallelism:

(d) “The man knew the solution to the problem was incorrect”

In this case, the noun phrase “the solution...” can be locally interpreted as either the direct object, which turns out to be incorrect, or an embedded subject, which is ultimately proved correct. Indeed, other interpretations are possible, such as the beginning of a possessive noun phrase:

(e) “The man knew the solution's discoverer was clever”

Theoretically, due to the recursive nature of the possessive construction, there are an infinite number of possible partial syntactic analyses for the fragment “The man knew the solution...”. Given that the human mind is finite, there is no way for it to simultaneously entertain an infinite number of partial parses.

In dealing with the problem of local ambiguity, we can therefore conclude that the human parser pursues either a single, serial analysis, or a bounded number of parallel analyses. In either case, we assume there is one analysis, which is “favoured” by the parser. For the remainder of this paper we will assume the simpler serial model. We believe, however, that what follows could be equally well applied to a bounded, ranked parallel mechanism.

Given that the general goal of the parser is to recover the most probable interpretation for the sentence as a whole, what strategy for resolving local ambiguities will best achieve this? The sentence processing literature has posited numerous parsing strategies, such as Frazier's (1979) Minimal Attachment and Late Closure principle. Such strategies, however, are typically motivated by the desire to minimise computational complexity or memory load, rather than obtaining a likely parse. Indeed, Hindle & Rooth (1993) demonstrate that for prepositional phrase attachment ambiguities, Minimal Attachment will make the incorrect attachment decision in more than 60% of instances (based on corpus findings).

Another strategy, which might be naturally assumed in the context of the present discussion, would be to adopt the most likely analysis at each local point of ambiguity. This suggestion is approximately what has been proposed under the heading of “constraint-based” theories of parsing, in the context of a parallel, competitive-activation architecture, . We argue, however, that while superficially appealing, this approach is not optimal for the resolution of local ambiguity.

Constraint-based theories are close to parallel likelihood accounts. In such models, parsing preferences are based on the simultaneous interaction of multiple constraints (MacDonald, 1994; MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, 1996; Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell, Tanenhaus, & Kello, 1993; cf. Taraban & McClelland, 1988; Tyler & Marslen-Wilson, 1977). These constraints relate to any properties of the encountered sentence that may influence its continuation, including subcategorisation preferences, other syntactic cues, the meaning of the fragment, the nature of the discourse context, and prosody or punctuation. Such models are unrestricted, in that all potentially relevant sources of information can be employed during initial processing (Pickering, 1997; Traxler & Pickering, 1996). Data about these sources of information can be estimated from corpus counts or obtained from production tasks where participants complete sentences from the point of ambiguity onwards. For instance, a few participants might complete “The man realised...” using a noun (e.g., “his goals”), but most will use a complement clause (e.g., “...his goals were unattainable”); and most participants might complete “The man realised his...” using a noun plus verb phrase on a complement-clause analysis (e.g., “...goals were unattainable”). From this, such models would assume that the parser would prefer the complement-clause analysis to the object analysis (e.g., Trueswell et al., 1993), and would foreground it by “his” at the latest. This is as predicted by likelihood accounts. However, recent constraint-based models sometimes assume that the parser pays attention to broader classes in making decisions (e.g., MacDonald, 1994; MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, 1996; Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell, Tanenhaus, & Kello, 1993). It might be noted that more verbs are transitive than intransitive, say, and thus support a transitive analysis of a fairly rare intransitive-preference verb. Unless the verb were extremely rare (with the data about preferences being unreliable), such a heuristic would go against likelihood. Constraint-based models therefore approximate to likelihood, but may diverge from it in some cases.

Formulation of the problem

We assume that the specific goal of the parser is to maximize the probability of recovering the parse of the input as generated by the speaker.[13] This specific goal is more useful than a general goal such as obtaining as much information as possible from the input, because the role of the parser is not to discover as much information, of whatever kind, as possible, but is to discover one particular piece of information, the correct structure of the input being analysed, so that this information can be used in building up an appropriate semantic representation of the input. Of course, the larger goal of the language processing system may be inquiry into the intentions of speakers, the state of the world, or even (according to a utility-based viewpoint) how to improve their choices of actions. But such larger goals have such a complex relationship to parsing a sentence (fundamentally because syntactic properties of sentences are not directly related to their semantic and pragmatic properties), it seems unlikely that such larger goals can be usefully related to the problem of choosing between alternative parses.

[13] Of course, the speaker will presumably not consciously intend any syntactic analysis, only the message conveyed. But presumably the language processing system of the speaker will have assigned a syntactic analysis to the sentence produced, or at least some representation (e.g., a semantic representation) which determines the correct global syntactic analysis. Note that there are some syntactic ambiguities which have no semantic consequences, but that these are not relevant here.

We shall assume that the performance of the parser on a given input can be classified unambiguously as either a success (the globally correct parse is recovered) or a failure (it is not). But this is clearly a significant oversimplification, because parsing may succeed to a greater or lesser degree. At one extreme it might lead to a structure for the sentence which is incorrect in some minor detail; at the other extreme it might lead to total breakdown early in the sentence, so that no useful information can be used to constrain semantic processes. The simplifying assumption that parses either succeed or fail provides, however, the starting point for a rich and tractable rational analysis. We hope that future research will consider how the analysis of a notion of graded success in parsing might be incorporated into this kind of account.

Processing assumptions

The task of resolving local ambiguity in parsing must proceed in the face of the two types of informational limitation--limited data, and limited computational resources with which to analyse that data--that we described in Part I. The limitation on data is inherent in the problem--it is this that means the parser faces a problem of local ambiguity; but the problem of limited data cannot be directly addressed by the parser, because the source of data (in speech understanding) is outside its control[14]. By contrast, the parser does have control over how it allocates computational resources in analysing the current input, so that it has the best chance possible of successfully navigating through the series of local ambiguities it faces under the severe real-time constraints of parsing fast speech. Therefore, it seems appropriate to setting the rational analysis of parsing within the framework of optimal computational resource allocation, rather than optimal data selection.

Any problem involving the allocation of computational resources must begin from some assumptions about the nature of the resources to be allocated. Fortunately, these assumptions can be very general constraints on the structure of the language processor, rather than requiring a fully detailed theory of language processing. First, we must ask: Is the parser serial, parallel, or somewhere in between? This is, of course, a central and unresolved issue in the study of human sentence processing, as we discussed above. In developing the analysis here, we choose to assume that the parser has a strict serial architecture, because this is the simplest, and most extreme, form of resource limitation in this context. We suspect that the conclusions from this analysis carry over, to some extent,

[14]In reading, by contrast, the parser has the opportunity to influence the way in which language input is sampled—e.g., by influencing fixation times, or by triggering regressive eye-movements. Of course, this control may not be exercised directly (which might appear unlikely according to a strong modularity viewpoint, e.g., Fodor, 1983), but instead by later processes which depend on the output of the parser.

when the assumption of strict seriality is weakened, to allow a modest degree of parallelism--but investigating this question rigorously remains a topic for future research.

A serial parser must choose a single option when it reaches a local ambiguity. Later in the sentence, it may become clear that the choice was incorrect. The parser must then “backtrack” to revise its previous choice. This choice, too, may be incorrect and subsequently require revision. Eventually, the parser makes the correct choice and further backtracking is not required. Eventually, if all goes well, the parser reaches the globally correct parse for the sentence. For simplicity, we shall assume that global syntactic ambiguity of the entire sentence is sufficiently rare that it can be ignored, so that a globally consistent parse, when found, will be the correct parse.

With limitless time and memory, and an adequate knowledge of the structure of the language, the parser could eventually always find a globally consistent parse for any grammatical sentence, simply by exhaustively searching through the space of possible sequences of parsing choices. But in practice, the parser has limited resources, and may fail. We assume that the principal source of failure in parsing occurs in abandoning one option, and attempting to backtrack to another--if irretrievable failure occurs at this stage, we say that the parser has “crashed”. We assume that the probability of backtracking successfully is determined by two independent factors: the difficulty of backtracking out of the current hypothesis; and then the difficulty of setting up the second parse. The independence assumptions means that the probability of successfully switching from parse1 to parse2 is the product of the probability of “escaping” from parse1, multiplied by the probability of setting up parse2.[15] To minimize the probability of crashing, the parser must aim to find the correct hypothesis as rapidly as possible--because every time it follows up a false lead which must be rejected, there is a possibility that it will crash. Notice that this is a very simple model of the limitations on the parser--it captures only the fact that back-tracking is error-prone. A more realistic account might also take into account the fact that the parser must work with an input which arrives in real-time, and that excessive backtracking may mean that the parser falls irretrievably behind the input which it is parsing.

We have so far noted that the parser may fail by crashing. But the need to minimize the amount of back-tracking to reduce the probability of crashing introduces another possible source of error: incorrectly rejecting the correct parse. Consider the common situation in which the parser makes a choice at a local ambiguity which comes to seem unlikely (although not impossible) given later context. If the parse is pursued, then it is

[15]It is possible that these processes are not strictly independent - for example, it is possible that difficulties in escaping from parse1 may lead to increased difficulties in setting up parse2 (see, for example, Pritchett, 1992).

likely that back-tracking will be required, with the associated risk of crashing. But if the parse is not pursued, then there is the risk that this parse was correct after all, and the possibility of a correct parse has thereby been missed[16]. An optimal choice of parsing order requires finding the correct balance between these two risks--of crashing during unnecessary search, and of mistakenly rejecting the correct parse.

To sum up, we assume that the parser chooses the order in which to consider alternative hypotheses at a point of local ambiguity to minimize the overall probability of a crash in parsing a sentence. Determining what choices the parser should make involves two steps. First, we show that, given certain assumptions, the <u>global</u> goal of reducing the probability of crashing throughout the entire sentence can be reduced to the <u>local</u> goal of reducing the probability of crashing at each specific local ambiguity. We then show how this local goal can be achieved.

<u>Step 1: From global to local maximization.</u>

Suppose that there is a single possible global parse, and that there are n local ambiguities, L_i, in the sentence. The correct parse will be reached if each of the n local ambiguities can be dealt with successfully (i.e., the correct hypothesis is adopted at that local ambiguity, perhaps after some backtracking).

$$\begin{aligned} P(\text{successful global parse}) &= P(\text{success at } L_1) \times P(\text{success at } L_2) \times \ldots \times P(\text{success at } L_n) \\ &= \prod_{i=1}^{n} P(\text{success at } L_i) \end{aligned} \tag{14}$$

Note that this independence assumption may be a significant simplification. For example, it is quite possible that if the local ambiguity L_i was only resolved with extreme difficulty, then the parser will be more likely to crash on the next local ambiguity, L_i. But we ignore such factors here and assuming that the chance of success at each local ambiguity is completely independent of the previous ones. Given independence, the goal of maximizing the probability of achieving the globally successful parse translates into the task of trying to maximize each element of the product--that is, at each local ambiguity L_i aiming to maximize the chance of choosing the correct parse at that ambiguity (where the “correct” local parse is simply determined by the single globally successful parse).

[16] We assume for simplicity that the parser cannot return to options that it has already rejected, although it would also be interesting to considering the analysis where this assumption is not made.

This means that the goal of maximizing the probability of a successful global parse can be translated into a sequence of independent local goals: maximizing the probability of the correct local parse, at each local ambiguity. That is, the parser must maximize:

$$P(\text{success at } L_i) \tag{15}$$

This will be the goal in all the analysis below.

Step 2: Locally optimal parse selection

We now consider how the parser should choose how to order its parses at a local ambiguity in order to maximize the probability that is gets through that ambiguity successfully (i.e., consistent with some globally possible parse). We call these parses "hypotheses," and label them $H_1 \ldots H_n$

An obvious suggestion is that we choose the hypotheses purely on the basis of the probability that they are correct, conditional on all the relevant information that can be accessed. We shall call these probabilities priors, because they are prior to assessing any of the hypotheses against subsequent context, and they are denoted $P(H_i)$.

We can assume, without loss of generality, that the hypotheses are ordered in some specific way, say: $H_1 \ldots H_n$: i.e. that H_1 is chosen first, and if this is rejected H_2 is chosen, and so on[17]. We will consider the probability that the parser eventually selects a correct hypothesis from these, given this order. We will then consider how the order should be optimized, to maximize the probability of success.

Suppose the hypothesis H_j is true. Then the probability of settling on it is the product of the probability of "escaping" from all the hypotheses which were considered previously, the $H_1 \ldots H_{j-1}$. This must be multiplied by the probability of successfully settling on H_j, which requires that the hypothesis H_j can be "set up" successfully, and then that the processor "sticks" with, rather than spuriously rejects, H_j.

For each of the $H_1 \ldots H_{j-1}$, the probability of escaping successfully from an erroneous hypothesis H_i successfully setting up, and rejecting, the hypothesis without crashing (assuming these are independent) is:

$$P(\text{escape } H_i) = P(\text{Set\_up } H_i | \overline{H_i}) P(\text{Reject } H_i | \overline{H_i}) \tag{16}$$

---

[17] Notice that by assuming a fixed order, we are ruling out the possibility that the nature of the failure on one hypothesis may determine the next hypothesis to try. This is, of course, only of importance in cases where there are more than two possible structures to be considered. To our knowledge, this possibility has not been proposed by other theorists, and we ignore it here.

Note that these probabilities are conditional on the fact that H_i is false, in symbols, $\overline{H_i}$. To finally settle on the correct hypothesis H_j requires successfully escaping from each of the *j*-1 previous hypotheses, which has probability:

$$\prod_{i<j} P(\text{escape } H_i) \tag{17}$$

Having "escaped" the preceding incorrect hypotheses, we consider the probability of settling on the correct hypothesis. For this to occur, the correct hypothesis must be set up correctly, and then not spuriously rejected. Again assuming these processes are independent, this gives:

$$P(\text{settle } H_i) = P(\text{Set\_up } H_i | H_i)\left(1 - P(\text{Reject } H_i | H_i)\right) \tag{18}$$

Combing (17) and (18), we conclude that, if the correct hypothesis is H_j the probability of choosing it is:

$$P(\text{settle } H_j)\prod_{i<j} P(\text{escape } H_i) \tag{19}$$

So far, we have assumed that the correct hypothesis is H_j. But, of course, at the point of local ambiguity, the correct hypothesis may be any of $H_1 \ldots H_n$, with priors $P(H_1) \ldots P(H_n)$. So the probability of settling on the correct hypothesis is a sum of terms of the form (19), weighted by the prior for each hypothesis:

$$P(\text{succeed at local ambiguity}) = \sum_{j=1}^{n}\left(P(H_j)P(\text{settle } H_j)\prod_{i<j} P(\text{escape } H_i)\right) \tag{20}$$

This expression assumes that we consider the hypotheses in the order $H_1 \ldots H_n$ We intend to maximize (20) by optimizing the order in which the hypotheses are considered.

The argument proceeds in two parts. The first part compares just two orders: $H_1 \ldots H_k H_{k+1} \ldots H_n$ and $H_1 \ldots H_{k+1} H_k \ldots H_n$, and shows that there is a simple rule for determining which order allows a higher value of (20). The second part shows that this rule determines the optimal order in which the hypotheses should be considered.

1: Comparing two similar orders. We denote the sequence $H_1 \ldots H_{k+1} H_k \ldots H_n$ by $G_1 \ldots G_k G_{k+1} \ldots G_n$, where $G_k = H_{k+1}$, $G_{k+1} = H_k$ and otherwise $G_i = H_i$. The probability of successfully ending with the correct hypothesis with the order of $\underline{H_{k+1}}$ and $\underline{H_k}$ swapped mirrors the expression (20):

$$P_G(\text{succeed at local ambiguity}) = \sum_{j=1}^{n} \left(P(G_j) P(\text{settle } G_j) \prod_{i<j} P(\text{escape } G_i) \right) \tag{21}$$

Swapping H_k and H_{k+1} improves the probability of successfully parsing the local ambiguity if and only if the expression in (21) is greater than (20). The derivation in Appendix A shows that this holds if and only if:

$$P(H_{k+1}) \times P(\text{settle } H_{k+1}) \times \frac{1}{1 - P(\text{escape } H_{k+1})} > P(H_k) \times P(\text{settle } H_k) \times \frac{1}{1 - P(\text{escape } H_k)} \tag{22}$$

This means that swapping H_k and H_{k+1} improves the probability of successfully parsing the local ambiguity if and only if condition (22) holds.

2: Establishing an optimal order. We have now provided a way of deciding between two orders of the form $H_1 \ldots H_k H_{k+1} \ldots H_n$ and $H_1 \ldots H_{k+1} H_k \ldots H_n$. Condition (22) shows that the choice is determined purely by properties of the hypotheses H_k and H_{k+1}, and is completely independent of all the other hypotheses, and the probabilities associated with them. This means that we can write (22) as

$$f(H_{k+1}) > f(H_k) \tag{23}$$

where $f(H_i) = P(H_i) \times P(\text{settle } H_i) \times \dfrac{1}{1 - P(\text{escape } H_i)}$.

Appendix B proves that the optimal order in which to test hypotheses is in descending order of $f(H_i)$ (where ties can be broken arbitrarily).

Thus, we have the conclusion that the parser should first consider the hypothesis with the highest value of $f(H_i)$. This condition has a simple intuitive meaning. The first term of $f(H_i)$ indicates that, other things being equal, the hypothesis with the highest prior should be tested first. The second term indicates that, other things being equal, hypotheses which, if correct, are easy to settle on, are to be tested earlier. The third term indicates that, other things being equal, the hypothesis from which it is easiest to escape when it is not correct should be preferred. This rational analysis thereby provides a strong set of predictions about parsing preferences, assuming that parsing preferences are at least approximately optimal. Testing the empirical adequacy of this rational analysis will require both empirical data concerning parsing, but also analysis of natural language corpora (to obtain, for example, appropriate estimates of priors for different parsing hypotheses in the context of particular types of local ambiguities).

We stress that this rational analysis of parsing preferences is one of many possible rational analyses that could be devised. As Anderson (1990, 1991) stresses, finding a good rational analysis, like finding a good scientific theory of any kind, involves an iterative interaction of theory and data. We are currently obtaining relevant empirical evidence, but this is beyond the scope of this chapter, and we expect that this will require adjustment and elaboration of this account. For the present, we hope that this analysis will provide a useful starting point from which the iteration between theory and data can begin in the case of parsing, and that it serves more generally as an illustration of how a rational analysis of a cognitive process involving inquiry can be developed. Specifically, we hope that it usefully illustrates the simplifications that may be involved in moving from a general rational framework to a specific and tractable rational analysis of cognitive performance.

Conclusion

The rational analysis of inquiry is a large, and relatively unexplored, area of future research. As we have seen, a wide range of cognitive processes involve the distribution of resources, either in information gathering or information processing, so that inquiry is as effective as possible. We have also seen that there are formal frameworks in terms of which such analyses can be provided. So, from a normative point of view, the prospects for rational analysis of inquiry seem bright.

But rational analysis must be normatively justified <u>and</u> descriptively adequate. Thus, a crucial question for future research concerns the degree to which the predictions of rational analysis of cognitive processes involving inquiry can predict empirical data. We have noted already that Anderson's rational analysis of problem solving and Oaksford and Chater's (1994) account of the selection task have captured a wide range of empirical data.

In addition, however, we note that there is an fundamental reason to believe that the rational analysis of inquiry should, ultimately, be possible. Our cognitive processes are remarkably effective information processors--far more effective than any artificial system that we can devise at dealing with the uncertainty and complexity of the real world. Inquiry is so central to cognition that it seems likely that the processes of guiding inquiry must, similarly, be highly effective. Such success seems to require following, to some level of approximation, some normatively justified principles for guiding inquiry--otherwise this success is inexplicable. If this is correct, the rational analysis of cognitive processes involving inquiry must be possible: there must be some normatively justified account to which the cognitive system approximates. Arguments for the mere existence of rational analyses, are, of course, only partially reassuring. The crucial issue is whether it is possible to devise rational analyses which are normatively and descriptively adequate. We hope that this chapter has shown that this is indeed sufficiently likely to justify future research.

References

Anderson, J. R. (1990). The adaptive character of thought. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Anderson, J. R. (1991). Is human cognition adaptive? Behavioral and Brain Sciences, 14, 471-517.

Arrow, K.J., Colombatto, E., Perlman, M. & Schmidt, C. (1996). The rational foundations of economic behavior. London: Macmillan.

Berger, J. O. (1985). Statistical decision theory and Bayesian analysis. New York: Springer-Verlag.

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

Colman, A. (this volume). Rationality assumptions and game theory and the backward induction paradox.

de Finetti, B. (1972). Probability, induction and statistics. New York: John Wiley.

deGroot, M. H. (1970). Optimal statistical decisions. New York: McGraw-Hill.

Earman, J. (1992). Bayes or bust? Cambridge, MA: MIT Press.

Elster, J. (1983). Sour grapes. Cambridge: Cambridge University Press.

Elster, J. (1989). Solomonic judgements. Cambridge: Cambridge University Press.

Evans, J. St. B. T. & Over, D. E. (1996). Rationality in the selection task: Epistemic utility versus uncertainty reduction. Psychological Review, 103, 356-363.

Frazier, L. (1979). On comprehending sentences: Syntactic parsing strategies. Ph.D. Thesis, University of Connecticut. West Bend, IN: Indiana University Linguistics Club.

Fodor, J. A. (1983). Modularity of mind. Cambridge, MA: MIT Press.

Gibson, J. J. (1979). The ecological approach to visual perception. Boston: Houghton-Mifflin.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilisitic mental models: A Brunswickian theory of confidence. Psychological Review, 98, 506-528.

Good, I. J. (1966). A derivation of the probabilistic explication of information. Journal of the Royal Statistical Society, Series B, 28, 578-581.

Hindle, D. & Rooth, M. (1993). Structural Ambiguity and Lexical Relations, Computational Linguistics, 19, 103-120.

Howson, C., & Urbach, P. (1989). Scientific reasoning: The Bayesian approach. La Salle, IL: Open Court.

Kahneman, D., (1996). Challenges to the rationality assumption. In K.J. Arrow, E. Colombatto, M. Perlman & C. Schmidt (Eds.) (1996). The rational foundations of economic behavior. London: Macmillan.

Kahneman, D., Slovic, P. & Tversky, A. (Eds.) (1982). Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press.

Kamil, A. C., Krebs, J. R. & Pulliam, H. R. (Eds.) (1987). Foraging theory. New York: Plenum Press.

Keynes, J. M. (1921). A treatise on probability. London: Macmillan.

Lettvin, J. Y., Maturana, W. S., McCullough, W. & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. Proceedings of the Institute of Radio Engineers, 47, 1940-1951.

Lindley, D. V. (1956). On a measure of the information provided by an experiment. Annals of Mathematical Statistics, 27, 986-1005.

Lindley, D. V. (1971). Bayesian statistics: A Review. Philadelphia: Society for Industrial and Applied Mathematics.

MacDonald, M.C. (1994). Probabilistic constraints and syntactic ambiguity resolution. Language and Cognitive Processes, 9, 157-201.

MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. Psychological Review, 101, 676-703.

Mackay, D. J. C. (1992). Information-based objective functions for active data selection. Neural Computation, 4, 590-604.

Manktelow, K. & Over, D. E. (1987). Reasoning and rationality. Mind and Language, 2, 199-219.

McNamara, J. M. & Houston, A. I. (1992). Risk-sensitive foraging: A review of the theory. Bulletin of Mathematical Biology, 54, 355-378.

Marr, D. (1982). Vision. San Francisco, CA: Freeman.

Norman, D. A. & Bobrow, D. G. (1975). On data-limited and resource-limited processes. Cognitive Psychology, 7, 44-64.

Oaksford, M. (1989). Cognition and inquiry: The pragmatics of conditional reasoning. Unpublished doctoral dissertation, Centre for Cognitive Science, University of Edinburgh.

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). Theories of reasoning and the computational explanation of everyday inference. Thinking and Reasoning, 1, 121-152.

Oaksford, M., & Chater, N. (this volume). A revised rational analysis of the selection task: Exceptions and sequential sampling.

Oaksford, M., & Chater, N. (1995b). Information gain explains relevance which explains the selection task. Cognition, 57, 97-108.

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. Psychological Review, 103, 381-391.

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.

Over, D. E. & Jessop, A. (this volume). Rational analysis of causal conditionals and the selection task.

Pickering, M.J. (1997). Sentence comprehension as an adaptive process. In Garrod, S.C. & Pickering, M.J. (Eds.) Language processing. London: UCL Press.

Pritchett, B. (1992). Grammatical competence and parsing performance. Chicago: University of Chicago Press.

Sen, A, (1990). Rational behavior. In J. Eatwell, M. Milgate and P. Newman (Eds.), The New Palgrave: Utility and Probability. London: Macmillan.

Shannon, C. E. (1948). The mathematical theory of communication. Bell System Technical Journal, 27, 379-423, 623-656.

Simon, H. (1982). Models of bounded rationality. (2 Vols.). Cambridge, MA: MIT Press.

Stalnaker, R. (1984). Inquiry. Cambridge, MA: MIT Press.

Taraban, R., & McClelland, J.R. (1988). Constituent attachment and thematic role assignment in sentence processing: Influence of content-based expectations. Journal of Memory and Language, 27, 597-632.

Traxler, M. J. & Pickering, M. J. (1996). Case marking in the parsing of complement sentences: Evidence from eye movements. Quarterly Journal of Experimental Psychology, 49A, 991-1004.

Trueswell, J.C. (1996). The role of lexical frequency in syntactic ambiguity resolution. Journal of Memory and Language, 35, 566-585.

Trueswell, J., Tanenhaus, M.K., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. Journal of Memory and Language, 33, 285-318.

Trueswell, J., Tanenhaus, M.K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. Journal of Experimental Psychology: Learning, Memory and Cognition, 19, 528-553.

Tyler, L.K., & Marslen-Wilson, W.D. (1977). The on-line effects of semantic context on syntactic processing. Journal of Verbal Learning and Verbal Behavior, 16, 683-692.

von Neumann, J. & Morgenstern, O. (1944). Theory of games and economic behavior. Princeton: Princeton University Press.

Wald, A. (1947). Sequential analysis. New York: Wiley.

Wald, A. (1950). Statistical decision functions. New York: Wiley.

Wason, P. C. (1966). Reasoning. In B. Foss (ed.), New horizons in psychology, Harmondsworth, Middlesex: Penguin.

Wason, P. C. (1968). Reasoning about a rule. Quarterly Journal of Experimental Psychology, 20, 273-281.

*Young, R. (this volume)

Appendix A: Derivation involved in comparing two parsing orders

We denote the sequence $H_1 \ldots H_{k+1} H_k \ldots H_n$ by $G_1 \ldots G_k G_{k+1} \ldots G_n$, where $G_k = H_{k+1}$, $G_{k+1} = H_k$ and otherwise $G_i = H_i$. By the analysis in the text, swapping the order of H<u>k+1</u> and H<u>k</u> increases the probability of choosing the correct hypothesis if and only if (21) > (20). The following derivation shows that the relationship can be expressed in a simple form. We first expand (21) as follows:

$$= \sum_{j=1}^{k-1}\left(P(G_j)P(\text{settle } G_j)\prod_{i<j} P(\text{escape } G_i)\right)$$
$$+P(G_k)P(\text{settle } G_k)\prod_{i<k} P(\text{escape } G_i)$$
$$+P(G_{k+1})P(\text{settle } G_{k+1})\left(P(\text{escape } G_k)\prod_{i<k} P(\text{escape } G_i)\right)$$
$$+\sum_{j=k+2}^{n}\left(P(G_j)P(\text{settle } G_j)\left\{\left(\prod_{i=1}^{k-1} P(\text{escape } G_i)\right)P(\text{escape } G_k)P(\text{escape } G_{k+1})\left(\prod_{i=k+2}^{j} P(\text{escape } G_i)\right)\right\}\right)$$

$$= \sum_{j=1}^{k-1}\left(P(H_j)P(\text{settle } H_j)\prod_{i<j} P(\text{escape } H_i)\right)$$
$$+P(H_{k+1})P(\text{settle } H_{k+1})\prod_{i<k} P(\text{escape } H_i)$$
$$+P(H_k)P(\text{settle } H_k)\left(P(\text{escape } H_{k+1})\prod_{i<k} P(\text{escape } H_i)\right)$$
$$+\sum_{j=k+2}^{n}\left(P(H_j)P(\text{settle } H_j)\left\{\left(\prod_{i=1}^{k-1} P(\text{escape } H_i)\right)P(\text{escape } H_{k+1})P(\text{escape } H_k)\left(\prod_{i=k+2}^{j} P(\text{escape } H_i)\right)\right\}\right)$$

which gives:

$$
\begin{aligned}
&= \sum_{j=1}^{k-1} \left(P(H_j) P(\text{settle } H_j) \prod_{i<j} P(\text{escape } H_i) \right) \\
&+ P(H_{k+1}) P(\text{settle } H_{k+1}) \prod_{i<k} P(\text{escape } H_i) \\
&+ P(H_k) P(\text{settle } H_k) \left(P(\text{escape } H_{k+1}) \prod_{i<k} P(\text{escape } H_i) \right) \\
&+ \sum_{j=k+2}^{n} P(H_j) P(\text{settle } H_j) \prod_{i<j} P(\text{escape } H_i)
\end{aligned}
\tag{A1}
$$

For comparison with the original order <u>H</u>, we now rearrange (20) in a similar form:

$$
\begin{aligned}
&= \sum_{j=1}^{k-1} \left(P(H_j) P(\text{settle } H_j) \prod_{i<j} P(\text{escape } H_i) \right) \\
&+ P(H_k) P(\text{settle } H_k) \prod_{i<k} P(\text{escape } H_i) \\
&+ P(H_{k+1}) P(\text{settle } H_{k+1}) \left(P(\text{escape } H_k) \prod_{i<k} P(\text{escape } H_i) \right) \\
&+ \sum_{j=k+2}^{n} P(H_j) P(\text{settle } H_j) \prod_{i<j} P(\text{escape } H_i)
\end{aligned}
\tag{A2}
$$

Order <u>G</u> is better than order <u>H</u> if expression (A1) is greater than expression (A2). This will be true if and only if:

$$
\begin{aligned}
& P(H_{k+1}) P(\text{settle } H_{k+1}) \prod_{i<k} P(\text{escape } H_i) \\
&+ P(H_k) P(\text{settle } H_k) \left(P(\text{escape } H_{k+1}) \prod_{i<k} P(\text{escape } H_i) \right) \\
&> P(H_k) P(\text{settle } H_k) \prod_{i<k} P(\text{escape } H_i) \\
&+ P(H_{k+1}) P(\text{settle } H_{k+1}) \left(P(\text{escape } H_k) \prod_{i<k} P(\text{escape } H_i) \right)
\end{aligned}
\tag{A3}
$$

which is true if and only if

$$P(H_{k+1})P(\text{settle } H_{k+1}) + P(H_k)P(\text{settle } H_k)P(\text{escape } H_{k+1})$$
$$> P(H_k)P(\text{settle } H_k) + P(H_{k+1})P(\text{settle } H_{k+1})P(\text{escape } H_k)$$

$$P(H_{k+1})P(\text{settle } H_{k+1})(1 - P(\text{escape } H_k)) > P(H_k)P(\text{settle } H_k)(1 - P(\text{escape } H_{k+1}))$$

$$P(H_{k+1}) \times P(\text{settle } H_{k+1}) \times \frac{1}{1 - P(\text{escape } H_{k+1})} > P(H_k) \times P(\text{settle } H_k) \times \frac{1}{1 - P(\text{escape } H_k)} \qquad \text{(A4)}$$

This gives the result stated in the text as (22).

Appendix B: Proof that the optimal order is to choose hypotheses in descending order of $f(H_i)$

We first note that there is an order which is optimal (there may be more than one such order as ties are possible), because there are only a finite number of orders.

We then use Reductio Ad Absurdem. That is, we assume that an optimal order E is not in descending order of $f(H_i)$. This implies that there is at least one pair of adjacent hypotheses, H_m and H_n, where H_m is chosen before H_n, but where $f(H_n) > f(H_m)$.

Now consider the order D, which is the same as E, except that H_n is chosen before H_m. By Part I, D is better than E. But this means that E is not optimal. This contradicts our original assumption that E was optimal. Therefore the optimal order is to select hypotheses in descending order of $f(H_i)$. The completes the proof.