A Rational Analysis of the Selection Task as Optimal Data Selection

Mike Oaksford and Nick Chater

Human reasoning in hypothesis-testing tasks like Wason's (1966, 1968) selection task has been depicted as prone to systematic biases. However, performance on this task has been assessed against a now outmoded falsificationist philosophy of science. Therefore, the experimental data is reassessed in the light of a Bayesian model of optimal data selection in inductive hypothesis testing. The model provides a rational analysis (Anderson, 1990) of the selection task that fits well with people's performance on both abstract and thematic versions of the task. The model suggests that reasoning in these tasks may be rational rather than subject to systematic bias.

Over the past 30 years, results in the psychology of reasoning have raised doubts about human rationality. The assumption of human rationality has a long history. Aristotle took the capacity for rational thought to be the defining characteristic of human beings, the capacity that separated us from the animals. Descartes regarded the ability to use language and to reason as the hallmarks of the mental that separated it from the merely physical. Many contemporary philosophers of mind also appeal to a basic principle of rationality in accounting for everyday, folk psychological explanation whereby we explain each other's behavior in terms of our beliefs and desires (Cherniak, 1986; Cohen, 1981; Davidson, 1984; Dennett, 1987; but see Stich, 1990). These philosophers, both ancient and modern, share a common view of rationality: To be rational is to reason according to rules (Brown, 1989). Logic and mathematics provide the normative rules that tell us how we should reason. Rationality therefore seems to demand that the human cognitive system embodies the rules of logic and mathematics. However, results in the psychology of reasoning appear to show that people do not reason according to these rules. In both deductive (Evans, 1982, 1989; Johnson-Laird & Byrne, 1991; Wason & Johnson-Laird, 1972) and probabilistic reasoning (Tversky & Kahneman, 1974), people's performance appears biased when compared with the standards of logic and probability theory.

Recently, however, some psychologists and philosophers have offered a different account of what it is to be rational (Anderson,

Mike Oaksford, Department of Psychology, University of Wales, Bangor, Wales, United Kingdom; Nick Chater, Department of Psychology, University of Edinburgh, Edinburgh, Scotland.

Correspondence concerning this article should be addressed to Mike Oaksford, who is now at Department of Psychology, University of Warwick, Coventry CV4 7AL United Kingdom; or to Nick Chater, who is now at Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD United Kingdom. Electronic mail may be sent to mike@cogsci.ed.ac.uk.

1990; Evans, 1993; Stich, 1990). In particular, Anderson (1990) argued that we must distinguish *normative* from *adaptive* rationality. An organism's behavior is rational if it is optimally adapted to its environment, even if reasoning according to logical rules had no causal role in producing the behavior. Such optimality assumptions have become widespread in contemporary social and behavioral science, from economics (Simon, 1959) to optimal foraging theory (MacFarland, 1977; MacFarland & Houston, 1981). Moreover, Anderson has extended this approach to provide "rational analyses" of memory, categorization, and problem solving (Anderson, 1990, 1991a, 1991b; Anderson & Milson, 1989).

In this article, we apply this approach to Wason's (1966, 1968) selection task, which has raised more doubts over human rationality than any other psychological task (Cohen, 1981; Manktelow & Over, 1993; Stich, 1985, 1990). In the selection task, an experimenter presents subjects with four cards, each with a number on one side and a letter on the other, and a rule of the form *if p, then q*, for example, *if there is a vowel on one side (p), then there is an even number on the other side (q)*. The four cards show an *A* (*p* card), a *K* (*not-p* card), a *2* (*q* card)< and a *7* (*not-q* card; see Figure 1). Subjects have to select those cards that they must turn over to determine whether the rule is true or false. Logically, subjects should select only the *p* and *not-q* cards. However, only 4% of subjects make this response, other responses being far more common (*p* and *q* cards, 46%; *p* card only, 33%; *p*, *q*, and *not-q* cards (7%); and *p* and *not-q* cards, 4%; Johnson-Laird & Wason, 1970a).

The selection task is a laboratory version of the problem of choosing the best experiments to test scientific laws. Popper's (1959) method of falsification provides the standard normative account of this situation. Popper argued that, logically, experiments can only falsify general laws, they cannot confirm general laws. Hence, scientists should only conduct experiments that can falsify a general law. The selection task provides an opportunity to see whether people spontaneously adopt Popper's falsificationist strategy (Wason & Johnson-Laird, 1972). Logically, the only way to falsify the conditional rule *if p, then q* in the selection task is to look for cards with *p* on one side and *not-q* on the other. Only two visible card faces are potentially of this type: the *p* card and the *not-q* card. Hence, according to falsifi-
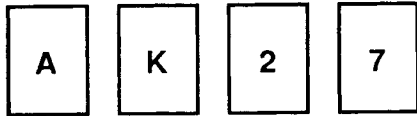
Figure 1. The four cards in the abstract version of Wason's (1966, 1968) selection task.

cation, subjects should choose only these two cards. However, in the selection task, as few as 4% of subjects make this card selection. This lack of fit between normative theory and behavior is responsible for the widespread doubts over human rationality we mentioned above.

Contemporary philosophers of science have rejected falsificationism as unfaithful to the history of science (Kuhn, 1962; Lakatos, 1970) and to be anyway unworkable (Churchland, 1986; Duhem, 1914/1954; Putnam, 1974; Quine, 1953). More recent accounts of scientific inference take a Bayesian probabilistic approach to confirmation (Earman, 1992; Horwich, 1982; Howson & Urbach, 1989). In particular, the Bayesian theory of optimal data selection (Federov, 1972; MacKay, 1992) offers a different account of how scientists should choose experiments that do not place an exclusive emphasis on falsification. Using this theory to develop a rational analysis of the selection task fits well with other rational analyses (e.g., Anderson 1990) that also use Bayesian methods. Our rational analysis will show that we can view behavior in the selection task as optimizing the expected amount of information gained by turning each card.

The purpose of a rational analysis is to show that behavior is optimally adapted to the environment. Good fits between a rational analysis and behavior indicate only that such an analysis provides an organizing framework for describing the behavior. Whether the behavior is rational depends on whether the rational analysis adequately characterizes the environment. Anderson (1990) used diffuse Bayesian prior distributions to model the environment. Although we do not use such distributions, we do make some assumptions about the environment that we do not justify until the Discussion section. In particular, we assume that the properties that figure in causal relations are rare in the environment. We call this the *rarity assumption*. We show that we can organize the data on the selection task on the assumption that subjects act as Bayesian optimal data selectors with rarity. In the Discussion section, we argue that the environment respects rarity, and hence we can view peoples' behavior on the selection task as adaptively rational.

The organization of this article is as follows. In the first section, we develop our rational analysis. In the following sections, we apply this analysis to a range of selection task data: the standard abstract results, the nonindependence of card selections (e.g., Pollard, 1985), the negations paradigm (e.g., Evans & Lynch, 1973), tasks that vary the probabilities of so-called fictional outcomes (Kirby, 1994), the therapy experiments (e.g., Wason, 1969), the reduced array selection task (RAST; e.g., Johnson-Laird & Wason, 1970b), and the thematic selection tasks (e.g., Cheng & Holyoak, 1985). Finally, we discuss the assumptions and implications of our account.

Rational Analysis

In this section, we first informally outline the problem of optimal data selection and how it applies to the selection task. We then present the Bayesian approach to optimal data selection. We then apply this account to derive a rational analysis of the selection task. Finally, we explore some general properties of the model's behavior.

Informal Outline

Optimal data selection involves choosing experiments to decide between rival hypotheses (Federov, 1972; Good, 1966; Hill & Hunter, 1969; Lindley, 1956; Luttrell, 1985; MacKay, 1992). For example, suppose that a metallurgist has various competing hypotheses about the underlying relationship between temperature and tensile strength. To decide between these hypotheses the metallurgist must choose new temperatures at which to test a metal's tensile strength. Intuitively, the most informative temperatures will be those where the hypotheses make divergent predictions (Platt, 1964). The Bayesian theory of optimal data selection formalizes these intuitions.

Everyday hypothesis testing also involves optimal data selection. Suppose that you are interested in the hypothesis that eating tripe makes people feel sick. In collecting evidence, should you ask known tripe eaters or tripe avoiders whether they feel sick? Should you ask people known to be, or not to be, sick whether they have eaten tripe? This case is analogous to the selection task. Logically, the hypothesis can be written as a conditional sentence, *if you eat tripe (p)*, then you feel sick *(q)*. The groups of people that you may investigate then correspond to the various visible card options, *p*, *not-p*, *q*, and *not-q*. In practice, who is available will influence decisions about who to investigate. The selection task abstracts away from this practical detail by presenting one example of each potential source of data. In terms of our everyday example, it is like coming across four people, one known to have eaten tripe, one known not to have eaten tripe, one known to feel sick, and one known not to feel sick. You must then judge which of these people you should question about how they feel or what they have eaten.

Let us consider informally what seems to be a rational selection of data in this situation. First, asking a person who has eaten tripe *(p)* is likely to be informative. If this person feels sick, then the hypothesis gains some credence; if not, then this evidence falsifies the hypothesis. Second, asking whether a tripe avoider *(not-p)* feels sick is futile, because the rule says nothing about how people feel if they have not eaten tripe. Third, asking whether a person who feels sick *(q)* has eaten tripe is worthwhile. The hypothesis will gain credence if they have eaten tripe, although if they have not, no conclusion appears to be forthcoming. Fourth, the person who is not feeling sick *(not-q)* is also worth questioning. If they have eaten tripe, this evidence falsifies the hypothesis. If they have not, no conclusion appears to be forthcoming. In this example, it seems that the *p* card is certain to be informative, the *not-p* card certainly will not be, and the *q* and *not-q* cards may or may not be informative. We now introduce the Bayesian approach to optimal data selection and show how to justify and extend these intuitions.

Bayesian Approach

We assume that the data that lead to the greatest expected reduction in uncertainty are the optimal data to select. We formalize uncertainty using Shannon–Wiener (Shannon & Weaver, 1949; Wiener, 1948) information. Given n mutually exclusive and exhaustive hypotheses (H_i), the uncertainty, $I(H_i)$, is

$$-\sum_{i=1}^{n} P(H_i)\log_2 P(H_i). \tag{1}$$

After obtaining some data D from performing an experiment (e.g., from turning a card in the selection task), we revise $P(H_i)$ to $P(H_i|D)$. Hence, the new uncertainty $I(H_i|D)$ is

$$-\sum_{i=1}^{n} P(H_i|D)\log_2 P(H_i|D). \tag{2}$$

We derive the $P(H_i|D)$ terms using Bayes' theorem:

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{\sum_{j=1}^{n} P(D|H_j)P(H_j)}, \tag{3}$$

which specifies the posterior probability of a hypothesis H_i given some data D in terms of the priors of each hypothesis H_j and the likelihoods of D given each H_j. The information gain, I_g, produced by a particular piece of data, is the reduction in uncertainty[1]:

$$I_g = I(H_i|D) - I(H_i). \tag{4}$$

Before collecting the data (i.e., turning a card), an inquirer does not know what the data outcome (the value of the hidden face) will be. Hence, it is not possible to calculate how informative the data will be. We therefore calculate the expected information gain, $E(I_g)$, taking into account all possible data outcomes (i.e., the two possible hidden faces of a card):

$$E(I_g) = E[I(H_i|D) - I(H_i)]. \tag{5}$$

Probability theory allows us to rewrite this as

$$E(I_g) = \left[\sum_{k=1}^{m} P(D_k)I(H_i|D)\right] - I(H_i), \tag{6}$$

where

$$P(D_k) = \sum_{j=1}^{n} P(H_j)P(D_k|H_j).$$

Thus, $E(I_g)$ is the uncertainty after performing the experiment, weighted by the probability of each possible experimental outcome, less the prior uncertainty. The greater $E(I_g)$, the more useful we predict an experiment is in deciding between rival hypotheses, H_i.

In this framework, hypotheses are not tested in isolation but in comparison to one another. This is for two reasons. First, uncertainty is only defined with respect to a set of alternative hypotheses (Equations 1 and 2). Second, the Bayesian calculations also require that such alternatives are defined (Equation 3). This comparative aspect of hypothesis testing is central to much recent philosophy of science (Earman, 1992; Lakatos, 1970; Platt, 1964).

Table 1
Probabilities for the Dependence Model, M_D and the Independence Model, M_I

	M_D		M_I	
	q	not-q	q	not-q
p	a	0	ab	$a(1-b)$
not-p	$(1-a)b$	$(1-a)(1-b)$	$(1-a)b$	$(1-a)(1-b)$

Note. a corresponds to the probability of p, $P(p)$, and b corresponds to the probability of q in the absence of p, $P(q|not\text{-}p)$.

Optimal Data Selection and the Selection Task

We now apply our Bayesian model of optimal data selection to the selection task. This involves specifying the alternative hypotheses that subjects must choose between and defining them probabilistically.

Accounting for the selection task requires considering just two hypotheses. One hypothesis, M_D, is that the dependency *if p, then q* holds. That is, if p occurs, then q must occur. We assume the simplest possible alternative hypothesis, M_I, that p and q are completely independent. Table 1 specifies the probabilities of each whole card, given each model. The left-hand side of the table specifies the probability, given M_D, of finding a p, q card, $P(p, q|M_D)$; a p, not-q card, $P(p, not\text{-}q|M_D)$; a not-p, q card, $P(not\text{-}p, q|M_D)$; and a not-p, not-q card, $P(not\text{-}p, not\text{-}q|M_D)$. Similarly, the right-hand side of Table 1 specifies the corresponding probabilities for M_I.

We set the probability of p, $P(p)$, to be the same in both models; that is, the row marginals are the same for M_D and M_I. Otherwise, observing p or *not-p* instances alone could provide evidence about whether the rule holds. In Table 1, we represent $P(p)$ by a. We also set the probability of q in the absence of p, $P(q|not\text{-}p)$, to be the same in both models. This reflects that a conditional rule makes a claim about the consequent only if the antecedent occurs but makes no claim otherwise (Quine, 1959). In Table 1, we represent $P(q|not\text{-}p)$ by b. The probability of q, $P(q)$, is not the same in M_D and M_I; that is, the column marginals are not the same. This is because the probability of q depends on whether the rule is true or false. If the rule holds, then the number of qs must be at least as great as the number of ps. There is no reason to assume this if p and q are independent.

We calculate expected information gain, $E(I_g)$, as follows. Equations 1 to 6 show how $E(I_g)$ can be derived using the prior probabilities of each hypothesis $P(H_i)$ and the conditional probabilities of each data outcome given each hypothesis, $P(D_k|H_i)$. We calculate the expected information gain associated with turning, say, the q card from the prior probabilities of the two hypotheses M_D and M_I and the probabilities of the two possible outcomes of turning the q card (i.e., a p or *not-p*).

We specify the priors by treating the probability of the inde-

---

[1] Oaksford (1989) provided a qualitative account of the selection task in terms of information gain in the situation-theoretic framework of Barwise and Perry (1983).
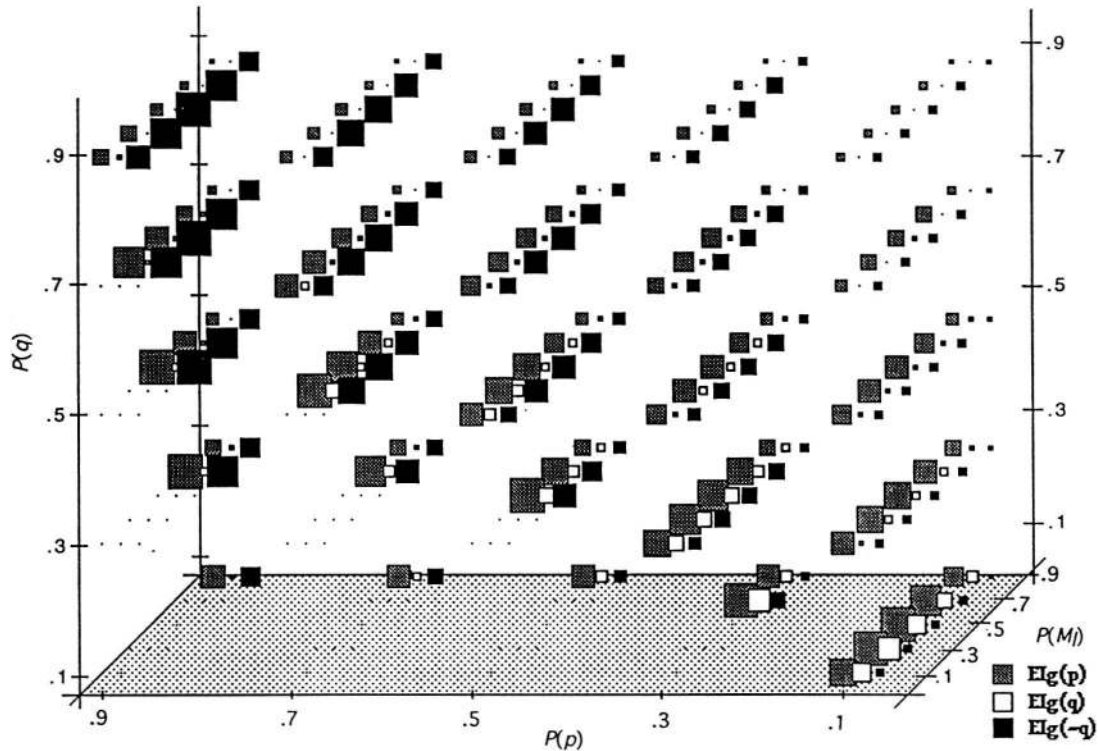
Figure 2. Model behavior. At each $[P(p), P(q), P(M_I)]$ coordinate, the three boxes represent the $E(I_g)$ values for the three cards. The area of the box is proportional to the $E(I_g)$ for the corresponding card. Coordinates where there are three dots indicate regions where the inequality $P(q) \geq P(q)P(M_D)$ is violated, and hence the probability values are inconsistent. M_I = independent model.

pendence model, $P(M_I)$, as a third parameter. The prior probability, of the dependence model $P(M_D)$, is therefore $1 - P(M_I)$.

To explain how we calculate the conditional probabilities, $P(D_k|H_i)$, let us first consider a slightly different set up from the standard selection task. Suppose instead of turning cards with one visible face you draw a card from a pack in ignorance of both sides. You can draw four possible cards: $p, q; p, not-q; not-p, q;$ or $not-p, not-q$. We can calculate $E(I_g)$s for this set up directly from the conditional probabilities of each of these cards given each model. These conditional probabilities are the cell values in Table 1. However, these are not the appropriate conditional probabilities to calculate $E(I_g)$ in the selection task. Calculations of the conditional probabilities, $P(D_k|H_i)$, must themselves be conditional on the visible face because the visible face is known. For example, if q is the visible face, then there are two possible outcomes if the card is turned: p or $not-p$. The probabilities of these outcomes, conditional on the hypothesis M_D and the visible q face, are $P(p|q, M_D)$ and $P(not-p|q, M_D)$, and similarly for M_I.[2] We derived these probabilities from Table 1 using identities such as

$$P(p|q, M_D) = \frac{P(p, q|M_D)}{P(p, q|M_D) + P(not\text{-}p, q|M_D)}, \quad (7)$$

$P(p, q|M_D) = a$, by the top left-hand cell in Table 1. $P(not\text{-}p, q|M_D) = (1 - a)b$, by the bottom left-hand cell in Table 1.

[2] There is an additional, although unimportant, complication, which concerns whether subjects can use the visible faces to tell between M_D and M_I. To see why this is possible, consider the following line of argument. If the experimenter drew the cards at random from a sample of cards and oriented these cards randomly to produce a range of visible faces, then in principle, the visible faces can contain useful information about which hypothesis is true. In particular, a visible q card is more likely if M_D is true than if M_I is true, and hence partially confirms M_D. Similarly, a visible $not\text{-}q$ face is less likely if M_D is true than if M_I is true, and hence partially confirms M_I. Because the p and $not\text{-}p$ cards are equally probable given either M_D or M_I, they do not confirm or disconfirm either model. As noted above, this reasoning depends on the assumption that the visible faces are randomly selected. In practice, of course, rather than making a random choice, the experimenter specifically chooses the cards so that there is precisely one visible face of each kind. Accordingly, the visible faces do not carry any information about which model is correct, and hence subjects cannot use the visible faces as evidence to tell between M_D and M_I. For this reason, we assume that subjects can derive no information from the visible faces. However, even if subjects did assume that the visible faces were generated by a random process, this would make little difference to our calculations. This is because taking visible cards into account would amount to revising the parameter that determines the prior probabilities of M_D and M_I, and, as we shall see, the predictions of our model turn out to be relatively independent of this parameter.

Hence, by Equation 7, $P(p|q, M_D) = a/[a + (1 - a)b]$. In this way, we computed the conditional probabilities required to calculate the $E(I_g)$s in the selection task in terms of $P(p)$ (i.e., a), b, and $P(M_D)$. Finally, we calculated b in terms of $P(M_D)$, $P(p)$, and the probability of q, $P(q)$:

$$b = \frac{P(q) - P(p)P(M_D)}{1 - P(p)P(M_D)}, \quad [P(q) \geq P(p)P(M_D)]. \quad (8)$$

In summary, we derive the $E(I_g)$s from the three probabilities $P(M_I)$ (recall that $P[M_D] = 1 - P[M_I]$), $P(p)$, and $P(q)$, which are the only free parameters in our model. The inequality $P(q) \geq P(p)P(M_D)$ must be respected if these parameters are to be consistent. This restriction arises because, if M_D holds, every p is associated with a q, and hence $P(q)$ must be at least as high as $P(p)$.

Model Behavior

We illustrate the behavior of the model in Figure 2. The three parameters, $P(p)$, $P(q)$, and $P(M_I)$ define a three-dimensional space. We calculated $E(I_g)$s for each card for five values (.1, .3, .5, .7, and .9) of each parameter. The not-p card does not appear in Figure 2 because its $E(I_g)$ value is always zero. At each coordinate, the three boxes represent the $E(I_g)$ values for the three cards. The area of the box is proportional to the $E(I_g)$ for the corresponding card.[3] Coordinates where there are three dots indicate regions where the inequality $P(q) \geq P(q)P(M_D)$ is violated, and hence the probability values are inconsistent. Figure 2 reveals the following pattern of expected informativeness for the four cards:

p card: Is informative insofar as $P(q)$ is low. It is largely independent of $P(p)$.[4]

q card: Is informative when $P(p)$ and $P(q)$ are both small.

not-q card: Is informative to the extent that $P(p)$ is large. It is independent of $P(q)$.

not-p card: Is not informative.

We highlight three aspects of the model's behavior. First, variation in $P(M_I)$ rescales the $E(I_g)$ values but does not change their order for the four cards. In consequence, the relative informational value of each card is insensitive to the priors (but see Footnote 4). Second, $E[I_g(not-p)]$ is always zero. This is consistent with the intuition that conditional rules make no assertions about what occurs if the antecedent is not satisfied. So, as we suggested above, consulting nontripe eaters cannot tell you whether eating tripe causes sickness. Third, when $P(p)$ and $P(q)$ are small (in the bottom right-hand corner of Figure 2), $E[I_g(q)]$ is greater than $E[I_g(not-q)]$. Figure 3 shows the entire region, R, (shown in black) where $E[I_g(q)] > E[I_g(not-q)]$, in more detail. Here, as in all subsequent analyses (unless explicitly stated otherwise), $P(M_I)$ is set to .5.[5] At all other values of $P(p)$ and $P(q)$, either $E[I_g(not-q)]$ is greater than $E[I_g(q)]$ or these values are undefined.

In modeling experimental data, we assume by default that $P(p)$ and $P(q)$ lie within R; that is, subjects treat p and q as rare. We refer to this assumption as the rarity assumption. Optimal data selection, together with the rarity assumption, will allow us to capture a wide range of experimental results. In the Discussion section, we consider whether the rarity assumption can it-
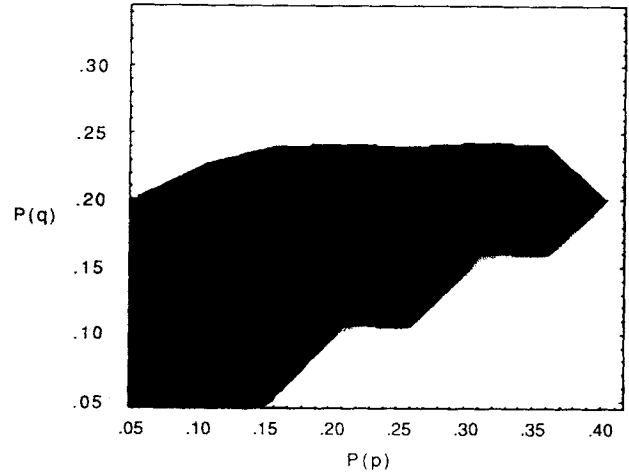

Figure 3. Plot of $P(p)$ against $P(q)$ with $P(M_I) = .5$, showing the region R (in black) where $E[I_g(q)] > E[I_g(not-q)]$.

self be rationally justified. But first, we use this framework to organize the data on the selection task.

Standard Abstract Results

Data

We described the standard abstract selection task in the introduction. Abstract tasks use unfamiliar content and contrast with thematic tasks that use familiar everyday contents. We discuss the thematic tasks later.

We conducted a meta-analysis (Wolf, 1986) of the abstract data that revealed the following ordering in individual card selection frequencies: $p > q > not-q > not-p$. Table 2 shows the results of the studies included in our meta-analysis.

We included all studies reporting individual card selection frequencies. We found 13 such studies reporting 34 standard abstract selection tasks, involving 845 subjects.[6] Table 2 shows the frequency

---

[3] We thank Mike Malloch for suggesting this method of visualization and for writing the software that generated this figure.

[4] However, when $P(M_I)$ is low (i.e., $P[M_D]$ is high), Equation 8 reveals that b can still be low. This leads to high values of $E[I_g(p)]$ when $P(M_I)$ is low, and $P(p)$ and $P(q)$ are high. See the top right-hand corner of Figure 3.

[5] Much theoretical debate concerning the validity of Bayesian statistics centers on how to assign prior probabilities, and there is a complex literature that considers how to do this (Earman, 1992; Jaynes, 1978; Skilling, 1989). When we have just two discrete hypotheses, that if p, then q holds and that p and q are completely independent, the most obvious prior, that each model has prior probability .5, is also the most theoretically justified (it corresponds to the maximum entropy prior and to Bayes' own principal of indifference [Bayes, 1763/1958]). We shall assume this prior in all subsequent analyses except when it is explicitly varied.

[6] By "affirmative" we mean that the task rule contained no negations. Below, we look at Evans's "negations paradigm," where negations are included in the task rules.

Table 2
Studies Reporting an Affirmative Abstract Version of the Selection Task

Study	Expt./condition	Card selection frequencies				
		p	q	*not-q*	*not-p*	N
Wason (1968)	Expt. 1/experimental	18	14	5	3	18
	Expt. 1/control	16	11	2	1	16
	Expt. 2	26	13	3	3	26
Evans & Lynch (1973)	Single expt.	21	12	8	2	24
Manktelow & Evans (1979)	Expt. 1/abstract	23	15	8	3	24
	Expt. 2/abstract	20	16	6	4	24
	Expt. 3/abstract	14	9	6	6	16
	Expt. 4/abstract	11	10	2	0	16
	Expt. 5/abstract	14	13	1	2	16
Griggs & Cox (1982)	Expt. 1/Trial 1	16	9	1	1	16
	Expt. 1/Trial 2	16	11	1	1	16
	Expt. 3/Trial 1	14	14	8	6	20
	Expt. 3/Trial 2	13	15	6	6	20
Griggs (1984)	Single expt./nonmem./T–F	21	13	2	4	25
	Single expt./nonmem./vio.	23	19	4	4	25
Chrostowski & Griggs (1985)	Single expt./nonmem./vio.	52	46	9	7	60
	Single expt./nonmem./T–F	58	47	5	4	60
Hoch & Tschirgi (1985)[a]	Single expt./bachelor's	22	15	10	6	25
Valentine (1985)	Single expt./AA	20	14	6	3	24
Yachanin (1986)	Expt. 2/widgit/vio.	20	16	4	3	20
	Expt. 2/widgit/test	19	14	6	4	20
Beattie & Baron (1988)	Expt. 1/four-card, +ve	17	10	1	1	18
	Expt. 2/four-card, +ve	16	10	8	1	16
	Expt. 3/four-card, +ve	16	10	2	0	16
Cosmides (1989)	Expts. 1 & 2	46	15	21	10	48
	Expts. 3 & 4	46	23	25	11	48
Girotto et al. (1992)	Expt. 1/arbitrary rule	20	14	6	4	24
	Expt. 2/arbitrary rule	23	13	8	7	24
	Expt. 3/arbitrary rule	19	10	13	7	24
	Expt. 4/arbitrary rule	16	12	5	2	20
Oaksford & Stenning (1992)	Expt. 2/abstract	19	15	5	6	24
	Expt. 3/colored shape	15	15	7	7	24
	Expt. 3/vowel–even	23	17	7	4	24
	Expt. 3/control	21	12	4	4	24
Totals		754	522	215	137	845
M proportion cards selected		.89	.62	.25	.16	
SD		.15	.12	.14	.09	

Note. For all studies using Evans's negations paradigm (Evans & Lynch, 1973), only the data for the affirmative rule are included. Studies were only included where individual card selection frequencies were reported or could be inferred from exhaustive reporting of card combinations. Expt. = Experiment; nonmem. = no memory cueing; T = true; F = false; vio. = violation condition; AA = affirmative antecedent and affirmative consequent condition; +ve = affirmative consequent condition.

[a] Only the bachelor's condition is included because the other two conditions (high school and master's) were not comparable to the subjects used in the remaining studies.

of individual card selections for each of these tasks and the average across studies. We performed a one-way analysis of variance taking task instance as the unit of analysis (see Glass, McGaw, & Smith [1981] for rationale), card type as the independent variable, and proportion of cards selected as the dependent variable.[7] This was highly significant, $F(3, 99) = 271.01$, $p < .0001$. Post hoc Tukey honestly significant difference tests revealed that each pairwise comparison between cards was significant at at least the .05 level. This provides strong evidence for the $p > q > $ *not-q* $ > $ *not-p* ordering in card selection frequencies.

Model

To model this ordering, we assume that by default subjects are operating in region R of Figure 3. For every such point in R

the expected information gain is ordered such that $E[I_g(p)] > E[I_g(q)] > E[I_g(not\text{-}q)] > E[I_g(not\text{-}p)]$. Mean $E(I_g)$ values sampled across R were $E[I_g(p)] = .76$; $E[I_g(q)] = .20$; $E[I_g(not\text{-}q)] = .09$; and $E[I_g(not\text{-}p)] = 0$.[8] This order mirrors the $p > q > $ *not-q* $ > $ *not-p* ordering in card selection frequencies. Therefore, card selection frequencies are monotonically related to expected information gain. This relationship suggests that subjects base

_____

[7] While advocating the use of Bayesian statistics in our model, we continue to use standard statistical tests in this article.

[8] We selected the points by laying a grid over Figure 3 with a mesh size of .025 on both axes. We then took the mean over all points that fell in R.

their card selections on the expected information gain of each card. We tested the prediction that the cards are ordered in this way using Page's L test for ordered alternatives (see Siegel & Castellan, 1988, pp. 184–188). As expected, this proved to be highly significant, $L(N = 34, k = 4) = 1,007.5, z_L = 9.36, p < .00001$.

Nonindependence of Card Selections

Data

Some studies have investigated whether card selections are statistically associated. In an analysis of just the q and not-q cards, Evans (1977) found that selection of these two cards was statistically independent. However, in a more detailed meta-analysis of three experiments, Pollard (1985) found consistent associations between card selections. He found that similarly valenced cards, that is, the p and q cards and the not-p and not-q cards, are positively associated, whereas selections of dissimilarly valenced cards—that is, p and not-p, p and not-q, q and not-p, and q and not-q—are negatively associated. Although these associations are not always statistically significant, their direction, positive or negative, is consistent across experiments.

We confirmed and extended these findings in a further meta-analysis. We took data from the studies analyzed by Pollard (1985) and five further studies, three from Oaksford and Stenning (1992; designated O & S2, O & S4, and O & S5) and two further unpublished control experiments (O & S1 and O & S3). All these experiments used task rules with negations varied in their antecedents and consequents, a manipulation that we discuss in the next section, Negations Paradigm Selection Task. Table 3 shows the results of our meta-analysis (see the rows labeled "AA" in Table 3). For the AA rule (affirmative antecedent, affirmative consequent) the following identities should be borne in mind: TA = p card; FA = not-p card; TC = q card; FC = not-q card (we explain the TA, FA, TC, and FC categories in the next section).

We performed the meta-analysis following Pollard (1985). We tested all six possible pairwise associations using Fisher's exact tests in the direction of the association present. We assigned a positive or negative z score to each result, setting z to 0 if the test yielded $p > .5$ in either direction (because this reveals a two-tailed probability of 1.0). We then calculated combined z estimates for each comparison and rule form using Stouffer's method (Wolf, 1986). Concentrating on the AA rule form, the combined estimates (see Table 3) were all significant (one-tailed) apart from the positive association between p (TA) and q (TC). The signs of the associations never reversed for any of the six pairs across all eight experiments. This was significant in one-tailed binomial tests (see Siegel & Castellan, 1988, pp. 38–44) for each of the six associations ($p < .005$).

Model

To model these associations we make three assumptions about how $E(I_g)$s map onto card selections. First, we assume that every card has some probability of being chosen because some subjects will simply not perform any, or an appropriate,

analysis. In particular, subjects will sometimes choose the not-p card, with $E[I_g(not$-$p)] = 0$. We account for this by adding a small fixed constant (.1) to the $E(I_g)$s for each card. Second, because of the four cards present in the selection task, we assume that card choice is a competitive matter. A card should have a greater chance of being chosen if it is less distinguishable from alternatives. One way to ensure that this happens for all four cards, including the not-p card, is to scale the $E(I_g)$s by the mean information available. We do this by dividing the derived score, $E(I_g) + .1$, for each card by the mean of this quantity for all four cards. We refer to this value as the scaled expected information gain, $SE(I_g)$. We assume that subjects choose cards as a monotonic function of their $SE(I_g)$ value. Third, a reasonable constraint on values for $P(p)$ and $P(q)$ is that $P(q) \geq P(p)$, otherwise the dependency model could not hold.

We sampled a variety of points corresponding to pairs of values for $P(p)$ and $P(q)$ at intervals of .025. The points satisfied the inequalities, $P(p) \leq .2$, $P(q) \leq .2$, $P(q) \geq P(p)$, and $P(q) \leq P(p) + .025$. The first two inequalities enforce the rarity assumption. The third inequality ensures that the dependency model can hold. The last inequality corresponds to the reasonable constraint that although the probability of q is greater than the probability of p, it is only marginally greater (Klayman & Ha, 1987). In the Discussion section, we discuss the justification of this constraint. We calculated $SE(I_g)$s for all four cards for each pair of $P(p)$ and $P(q)$ values (mean $SE[I_g]$s for each card appear in the AA row in Table 4). We then computed Spearman rank order correlation coefficients between the $SE(I_g)$s for all six card pairs. We used rank correlations because we assume only that card selection is a monotonic function of $SE(I_g)$. The results of these analyses appear in the AA column in Table 5. The $SE(I_g)$s for the similarly valenced cards are positively correlated, whereas the $SE(I_g)$s for the four dissimilarly valenced card comparisons are negatively correlated. This pattern of correlations is the same as that observed experimentally. The agreement in the sign of the correlation between model and data was significant in a one-tailed binomial test ($p < .025$).

This analysis applies only to the AA or purely affirmative rule. We now show how to extend this analysis to account for data from the negations paradigm selection task.

Negations Paradigm Selection Task

Data

In the negations paradigm selection task (Evans & Lynch, 1973), the antecedent and consequent of a rule can contain negated constituents (not-p, not-q). There are four possible conditional rules: the original $if p, then q$ (AA), together with $if p, then not q$ (affirmative antecedent, negative consequent [AN]); $if not$-$p, then q$ (negative antecedent, affirmative consequent [NA]); and $if not$-$p, then not$-q (negative antecedent, negative consequent [NN]). Each subject performs a selection task for each of these four rule types.

We have so far described the cards in the selection task in terms of p, q, not-p, and not-q. In the negations paradigm, cards are normally described in terms of whether they make the antecedent or consequent of the rule true or false. For example, con-

Table 3

Meta-Analysis of Nonindependence of Card Selections

Comparison	Rule form	Study from which z is derived								Comb. z	Overall comb. z
		E & L	M & E1	M & E2	O & S1	O & S2	O & S3	O & S4	O & S5		
TA vs. FA	AA	−0.70	−0.60	−1.84	−2.49	−1.55	−2.72	−1.55	−0.97	−4.39	
	NA	−2.50	−2.94	−2.69	−1.92	−2.38	−1.23	−4.00	−0.71	−6.49	−10.43
	AN	−1.38	0.00	−0.34	0.00	−0.98	−2.58	−2.52	−0.40	−2.90	
	NN	−2.06	−4.00	−4.00	−1.54	−2.72	−1.54	−4.00	−0.16	−7.08	
TA vs. TC	AA	0.00	0.00	0.87	0.37	0.00	0.00	0.00	0.43	0.59	
	NA	0.00	−1.40	0.00	0.00	0.00	0.29	0.80	0.00	−0.11	1.32
	AN	−2.68	−0.50	0.30	0.00	−1.27	−0.92	0.00	0.71	−1.54	
	NN	1.82	1.35	3.98	1.13	0.88	0.00	1.29	0.00	3.69	
TA vs. FC	AA	−0.67	−1.17	−2.28	−0.51	−0.16	0.00	−1.55	−0.55	−2.44	
	NA	0.89	1.59	1.67	−1.26	−0.88	0.00	1.13	0.00	1.11	−2.42
	AN	0.98	0.00	0.00	−1.55	0.05	0.00	−0.27	0.00	−0.28	
	NN	0.00	−2.46	−2.29	−0.61	−1.42	−0.67	−0.56	−1.13	−3.23	
FA vs. TC	AA	0.00	−0.12	−1.24	−2.29	0.00	0.00	−1.55	0.00	−1.84	
	NA	−0.15	2.09	0.32	−0.45	0.00	0.00	−1.29	−0.14	0.13	−1.21
	AN	0.89	1.59	1.67	2.58	0.00	0.16	−0.45	−0.71	2.03	
	NN	−0.89	−2.38	−2.91	0.00	0.00	0.00	−1.55	0.00	−2.73	
FA vs. FC	AA	1.27	0.87	2.66	2.19	1.20	0.00	0.00	1.56	3.45	
	NA	−0.72	−1.17	1.01	−1.74	−2.01	0.00	1.34	1.61	−0.59	2.55
	AN	−0.21	−0.65	−1.07	0.00	−0.05	0.00	0.00	0.00	−0.70	
	NN	0.00	2.41	2.65	1.56	0.00	0.00	0.00	1.73	2.95	
TC vs. FC	AA	−0.67	−1.17	−2.28	0.00	−1.29	−0.48	−0.37	−0.97	−2.56	
	NA	0.89	1.59	1.67	0.00	−0.82	0.00	−2.26	−2.10	−0.36	−4.89
	AN	0.98	0.00	0.00	−0.45	−2.05	0.00	−2.68	−1.85	−2.14	
	NN	0.00	−2.46	−2.29	−1.33	−2.85	0.00	−1.67	−2.75	−4.72	

Note. E & L = Evans and Lynch (1973); M & E1 and M & E2 = Manktelow and Evans (1979) Experiments 1 and 2; O & S2, O & S4, O & S5 correspond to Oaksford and Stenning's (1992) Experiment 3/control, Experiment 3/subject–predicate, and Experiment 3/vowel–even conditions, respectively; O & S1 and O & S3 are Oaksford and Stenning's two unpublished control experiments; Comb. = combined; TA = true antecedent; FA = false antecedent; TC = true consequent; FC = false consequent; AA = affirmative antecedent and affirmative consequent; NA = negative antecedent and affirmative consequent; AN = affirmative antecedent and negative consequent; NN = negative antecedent and negative consequent. z scores greater than 1.65 or less than −1.65 are significant at $p < .05$ (one-tailed).

Table 4

Mean SE(I$_g$)s for Each Card for Each of the Four Rule Types in the Negations Paradigm Selection Task

Rule type	Cards			
	TA	TC	FC	FA
AA	2.43 (19.67)	0.87 (13.83)	0.43 (6.33)	0.27 (4.00)
AN	2.31 (21.00)	0.55 (4.67)	0.69 (14.83)	0.44 (2.67)
NA	2.26 (16.00)	1.08 (15.83)	0.42 (8.67)	0.23 (1.23)
NN	2.12 (18.5)	0.62 (10.83)	0.95 (12.33)	0.31 (6.00)

Note. $p(N = 16) = .92$ ($p < .0001$). Figures in parentheses indicate the mean number of each card selected in the eight studies mentioned in the text. $SE(I_g)$ = scaled expected information gain; TA = true antecedent; TC = true consequent; FC = false consequent; FA = false antecedent; AA = affirmative antecedent and affirmative consequent; AN = affirmative antecedent and negative consequent; NA = negative antecedent and affirmative consequent; NN = negative antecedent and negative consequent.

sider the rule "if there is not an *A* on one side, then there is not a *2* on the other side." For this rule the *p* card (*A*) and the *q* (*2*) card are the false antecedent (FA) and false consequent (FC) cases, respectively. The *not-p* (*K*) and *not-q* (*7*) cards are the true antecedent (TA) and true consequent (TC) cases, respectively. The TA, FA, TC, and FC logical cases permit a uniform classification of the cards in the negations paradigm.

Using the negations paradigm, Evans and Lynch (1973) reported an effect that they called *matching bias*. Subjects tend to select the cards that are named in the rules, ignoring the negations. For example, consider the rule "if there is not an *A* on one side, then there is not a *2* on the other side," and the four cards showing an *A*, a *K*, a *2*, and a *7*. The matching response is to select the *A* (FA) and the *2* (FC) cards. The confirmatory response is to select the *K* (TA) and *7* (TC) cards, and the falsificatory response is to select the *K* (TA) and *2* (FC) cards. More recently it has become clear that matching occurs mainly for the consequent cards (Evans, 1989; Manktelow & Evans, 1979), whereas antecedent card selections accord with the logical case (subjects choose the TA card in preference to the FA card for all four rule forms).

Card orderings. As in the standard selection task, we can describe these data in terms of orderings over card selections. The ordering in card selection frequencies for affirmative consequent rules (*if p, then q; if not-p, then q*) is TA > TC > FC > FA. The ordering in card selection frequencies for negative consequent rules (*if p, then not-q; if not-p, then not-q*) is TA > FC > TC > FA. Both these orders are consistent with matching.

A meta-analysis confirmed these orderings. We analyzed the studies used in our meta-analysis of nonindependence of card selections, less O & S4 and O & S5. We omitted these studies because they included manipulations to counteract matching (see below). We performed a one-way analysis of variance for each rule type (AA, AN, NA, and NN) with card type as the independent variable and number of cards selected as the dependent variable. For the affirmative consequent rules these analyses were significant: AA, $F(3, 15) = 111.81, MS_e = 2.75, p < .0001$; NA, $F(3, 15) = 21.19, MS_e = 6.78, p < .0001$. For both rules the order of mean number of cards selected reflected the TA > TC > FC > FA order. Each pairwise comparison between

cards was significant at at least the .05 level in post hoc Neuman-Keuls tests except for the TA versus TC and FA versus FC comparisons for the NA rule. Similar analyses were also significant for the negated consequent rules: AN, $F(3, 15) = 100.64, MS_e = 4.45, p < .0001$; NN, $F(3, 15) = 11.64, MS_e = 13.69, p < .0005$. For both rules the order of mean number of cards selected reflected the TA > FC > TC > FA order. Each pairwise comparison between cards was significant at at least the .05 level in post hoc Neuman-Keuls tests except for the FA versus FC comparison for the AN rule and the TC versus FC comparison for the NN rule.

The mean frequencies of card selections for each rule type appear in parentheses in Table 4. This table reveals that the orderings in card selections are weaker for the negative antecedent rules, NA and NN. The NA rule also appears to cause comprehension problems. Subjects are significantly slower to comprehend this rule than each of the other three rule forms (Ormerod, Manktelow, & Jones, 1993). This result is unexpected because the standard finding has been that sentences containing negations are harder to comprehend (Fodor, Fodor, & Garrett, 1975). This would predict that a sentence with two negations should be harder to comprehend than similar sentences containing single negations. However, subjects comprehend the NA rule significantly more slowly than the NN and the AN rule.

Suppressing matching. Two experimental manipulations suppress matching and reestablish the TA > TC > FC > FA order for all four rule types in the negations paradigm. First, although Manktelow and Evans (1979) initially found matching even with thematic material, Reich and Ruth (1982) and Griggs and Cox (1983) found that matching disappeared when they used more realistic thematic material. Second, Oaksford and Stenning (1992) found that matching also disappears when the linguistic framing of the rules is more appropriate. They argued that rules such as "if there is not an *A* on one side, then there is a *2* on the other side" are ambiguous. The *not-A* card could be the *K*, or the *2* or the *7* cards. Oaksford and Stenning found that removing this ambiguity suppressed matching and reestablished the TA > TC > FC > FA order for all four rule types.

Nonindependence of card selections. Pollard (1985) also investigated associations between card selections for the negations

Table 5

Spearman Rank Order Correlation Coefficients Between SE(I_g)s for Each Card Pair for Each of the Four Rule Types in the Negations Paradigm Selection Task

Comparison	AA	AN	NA	NN
TA vs. TC	0.11 (0.59)	−0.47 (−1.54)	−0.85 (−0.11)	0.99 (3.69)
FA vs. FC	0.79 (3.45)	−0.70 (−0.09)	0.67 (−0.59)	0.92 (2.95)
TA vs. FC	−0.69 (−2.44)	−0.94 (−0.28)	0.54 (1.11)	−0.99 (−3.23)
TA vs. FA	−0.21 (−4.39)	−0.22 (−2.90)	0.97 (−6.49)	−0.92 (−7.08)
TC vs. FC	−0.64 (−2.56)	0.59 (−2.14)	−0.88 (−0.36)	−0.99 (−4.72)
TC vs. FA	−0.92 (−1.84)	−0.53 (2.03)	−0.93 (0.13)	−0.94 (−2.73)

Note. Phi coefficient $(r_\phi) = .64$ ($p < .025$; Fisher's exact test). Figures in parentheses indicate the combined z scores for each pairwise comparison taken from Table 3. $SE(I_g)$ = scaled expected information gain; AA = affirmative antecedent and affirmative consequent; AN = affirmative antecedent and negative consequent; NA = negative antecedent and affirmative consequent; NN = negative antecedent and negative consequent; TA = true antecedent; TC = true consequent; FA = false antecedent; FC = false consequent.

paradigm. His meta-analysis looked at associations between card cases (p, not-p, q, and not-q). He found positive associations for similarly valenced cards and negative associations for dissimilarly valenced cards, not only for the standard affirmative rule (AA) but also for the remaining three rules (AN: *if p, then not-q*; NA: *if not-p, then q*; and NN: *if not-p, then not-q*).

Our extended meta-analysis, shown in Table 3, shows these associations in terms of the logical case. The overall combined z score (final column in Table 3) for each pairwise comparison treats each rule in each study as a separate unit of analysis ($N = 32$). The final column of Table 3 shows that the signs of the combined z scores are positive for similarly valenced cards (TA vs. TC and FA vs. FC) and negative for dissimilarly valenced cards (TA vs. FA, TA vs. FC, TC vs. FA, and TC vs. FC). These results follow the pattern for the standard selection task, discussed in the section Nonindependence of Card Selections.

Although this finding is clear for the similarly valenced rules, AA and NN, the results for the dissimilarly valenced rules, AN and NA, are more ambiguous. We highlight this division in Table 5, where we summarize the combined z scores for each rule separately (see figures in parentheses).

Model

The key to understanding a variety of effects in the negations paradigm is the notion of a "contrast set" (Hampton, 1989; Oaksford & Stenning, 1992). Contrast sets provide the interpretations of negated constituents. For example, the interpretation of "Johnny didn't serve *tea*" (where the word in italics indicates the focus of a negation) is that he served a drink other than tea. In terms of set theory, the superordinate category "drinks" provides the universe of discourse. Contrast sets are plausible subsets of the complement in a universe of discourse. In our example, all other drinks less tea form the complement. When Johnny didn't serve tea it is more likely he served soft drinks rather than, for example, scotch on the rocks. Soft drinks is therefore the contrast set, that is, a plausible subset of the complement. Background knowledge may restrict the membership of the intended contrast set even further. So, in our example, coffee is perhaps the most likely single contrast set member.

This indicates that a negation rarely identifies the complement, that is, the whole set consisting of the superordinate category less the named constituent, as the intended contrast set. More commonly the intention is to identify much more restricted contrast sets. We now apply this behavior of contrast sets to the negations paradigm.

We have good reason to believe that $P(\text{TA})$ or $P(\text{TC})$ are greater when they are negated. This is because the class of things referred to by a constituent is generally smaller than the size of the contrast set defined by its negation. For example, there are many things John could have drunk when he didn't drink tea.[9] However, the intended contrast set is unlikely to be all drinks other than tea. We made the reasonable assumption that the probability of a contrast set does not exceed .5.[10] We therefore set the probabilities of unnegated constituents to vary between 0 and .2 as for the AA rule, and those of negated constituents to vary between .2 and .5.

Card orderings. We explained the AA rule above. For the other three rule types we sampled points at .025 intervals as in the section Nonindependence of Card Selections. As for the AA rule, we now outline our rationale for the region over which we sampled points for each rule form. For the AN rule, $P(\text{TA})$ is low but $P(\text{TC})$ is high. We therefore sampled points in the region that satisfied the inequalities, $P(p) \leq .2$ and $.2 \leq P(q) \leq .5$. For the NA rule, $P(\text{TA})$ is high but $P(\text{TC})$ is low. This rule therefore violates the inequality that $P(\text{TC}) \geq P(\text{TA})P(M_D)$ (see Equation 8); that is, this rule corresponds to the region of the parameter space in Figure 3 where $E(I_g)$ is undefined. Informally, an NA rule is like the hypothesis that all black things are ravens. This rule must be false because there are more black

---

[9] It is possible that the token frequencies of the constituent are so large that they might outweigh the token frequencies of all the other possible drinks (i.e., if tea is the most common drink). But if this is true for tea, it cannot be true for any other drink. Hence, in the domain of drinks, negations will still most commonly identify highly probable contrast classes.

[10] Note that this means that the probability of an event and its most plausible contrast set will rarely sum to 1, although the probability of an event and its complement must sum to 1.

things than ravens (i.e., the dependency model could not hold). So, to interpret an NA rule as having a chance of being true involves either revising $P(\text{TA})$ down or revising $P(\text{TC})$ up. Subjects appear to resolve the ambiguity by revising $P(\text{TA})$ down. The NA rule then leads to low $P(\text{TA})$ and $P(\text{TC})$ values. However, it is reasonable to assume that $P(\text{TA})$ is still greater than $P(\text{TC})$ because the antecedent is negated. We therefore sampled points that satisfied the inequalities, $P(p) \le .2$, $P(q) \le .2$, $P(q) < P(p)$, and $P(q) \ge P(p) - .025$. For the NN rule, both $P(\text{TA})$ and $P(\text{TC})$ are high. We modeled this rule by restricting the values of $P(q)$ and $P(p)$ to the range .2 to .5. We therefore sampled points from the region that satisfied the inequalities, $.2 \le P(p) \le .5$, $.2 \le P(q) \le .5$, $P(q) \ge P(p)$, and $P(q) \ge P(p) + .025$.

Mean $SE(I_g)$ values for points sampled across the four regions appear in Table 4. The mean number of cards selected in the eight studies used in our meta-analysis of the negations paradigm selection task appear in parentheses. As can be seen, there was a good fit between data and model, Spearman's $\rho(16) = .92$, $p < .0001$. Finally the need to resolve the ambiguity between revising $P(\text{TA})$ down or $P(\text{TC})$ up accounts for why subjects are significantly slower to comprehend the NA rule than each of the other three rule forms (Ormerod et al., 1993).

Suppressing matching. "Realistic" thematic content restores the TA > TC > FC > FA ordering in card selection frequencies for all rule forms (Reich & Ruth, 1982). We argue that this is because prior world knowledge restricts contrast sets to the most plausible member(s). For example, if Johnny didn't drink tea, then it is most likely that he drank coffee. In a context where drinking tea is a possibility, drinking, for example, scotch on the rocks is probably not. Relative to the class of drinks, tea and coffee are both rare (subject to the caveats in Footnote 9). Such examples show that familiar thematic material reduces contrast set size, thereby reestablishing rarity. Unrealistic thematic content fails to restore the TA > TC > FC > FA ordering because it (Manktelow & Evans, 1979) cannot engage prior knowledge to constrain contrast sets in this way.

The same reasoning explains the restoration of the TA > TC > FC > FA ordering for all rule forms in Oaksford and Stenning (1992). The intention in this task was for subjects to regard the *not-A* contrast set to consist of only the *K* card. However, the *K*, *2*, and *7* cards are all potential members of the *not-A* contrast set. To restrict this contrast set the materials had to indicate unambiguously that only other letters are potential contrast set members. Oaksford and Stenning (1992) used the original vowels and even numbers material used by Wason (1968). "Vowels" and its complement set "consonants" only apply to letters. In the context of the task, where *K* is the only consonant, it should therefore be clear that *K* is the only possible member of the *not-A* contrast set. The antecedent is therefore unambiguously about the *K* card, and so rarity again holds. This predicts the standard TA > TC > FC > FA ordering for all four rule forms as Oaksford and Stenning found.

Nonindependence of card selections. Even when negated constituents are used, similarly valenced cards show positive associations and dissimilarly valenced cards show negative associations. We modeled this behavior in the same way as the AA rule in the section Nonindependence of Card Selections. We calculated Spearman rank order correlation coefficients for

each card pair for the same sets of points used to calculate mean $SE(I_g)$ values in Table 4. These coefficients appear in Table 5 together with the combined z scores taken from Table 3 (in parentheses). We assessed the fit between data and model in terms of the direction (+/-) of association or correlation using the phi coefficient (Siegel & Castellan, 1988, pp. 232-235), which showed a significant fit between data and model ($r_\phi = .64$, $p < .025$). The fit was perfect for the AA and NN rules. For the AN and NA rules the fit was less good, although the model does capture some of the interesting differences between the AA-NN rules and the AN-NA rules. The poorest fit was for the NA rule where subjects must revise $P(\text{TA})$ down and where they experience comprehension problems. Such problems may lead to a residual matching tendency. The overall agreement between data and model was highly significant, as assessed by a one-tailed binomial test (19 agreements vs. 5 disagreements: $p < .005$).

Probabilities of Fictional Outcomes

Data

Recently, Kirby (1994) explicitly manipulated $P(p)$ in a selection task. To understand the rationale behind Kirby's studies, it is necessary to briefly outline his theoretical account of the selection task. This account uses the vocabulary of subjective expected utility and signal-detection theories. Kirby started from the falsificationist assumption that a subject's goal is to find p and *not-q* instances, which he called an *inconsistent outcome*. He then assumed that two factors determine performance. The first factor concerns the probability of an inconsistent outcome arising from a visible card face (C). So like our model, Kirby's took into account the probabilities of the "fictional" other sides of each card (fictional because the subjects never turn the cards). The second factor concerns the utilities associated with a card choice. There are four possibilities, which Kirby classified using signal-detection theory: a *hit*, that is, choosing a card with a hidden face that is inconsistent with the rule; a *miss*, that is, not choosing such a card; a *false alarm* (FA), that is, choosing a card with a hidden face that is consistent with the rule; and a *correct rejection*, (CR), that is, not choosing such a card.

Kirby (1994) proposed that a subject should choose a card when the posterior odds of an inconsistent outcome exceed a simple function of the utilities (see Equation 9). In deriving predictions for his Experiments 1 to 3, Kirby assumed that the utilities on the right-hand side of Equation 9 remain constant.

$$\frac{P(\text{inconsistent outcome present} \mid C)}{P(\text{inconsistent outcome absent} \mid C)} > \frac{U(\text{CR}) - U(\text{FA})}{U(\text{hit}) - U(\text{miss})} \quad (9)$$

On this analysis the q and *not-p* cards have probability 0 of yielding an inconsistent outcome. Therefore, as with other falsificationist accounts, Kirby (1994) predicted that subjects should never turn these cards. The interest of his account therefore centers on the p and *not-q* cards.

Kirby (1994) noted that in Wason's (1968) original rule, *if a card has a vowel on one side, then it has an even number on the other side*, the posterior odds of finding an inconsistent outcome

Table 6
SE(I_g)s for Each Card Using the P(p), P(q) values Used by Kirby (1994)
in His Experiments 1–3, With P(M_I) = .01

		Cards			
Condition	P(p) and P(q)	not-q	p	not-p	q
Experiment 1					
Small P set	1/1,001	0.746 (.49)	1.311 (.78)	0.745 (.29)	1.198 (.60)
Large P set	1,000/1,001	1.372 (.73)	0.911 (.55)	0.858 (.48)	0.858 (.41)
		$\rho(N = 8) = .87\ (p < .01)$			
Experiment 2					
Small P set	1/100	0.745 (.40)	1.332 (.91)	0.740 (.17)	1.182 (.50)
Medium P set	50/100	1.009 (.53)	1.337 (.80)	0.745 (.27)	0.909 (.57)
Large P set	90/100	1.248 (.61)	1.262 (.80)	0.737 (.38)	0.753 (.56)
		$\rho(N = 12) = .90\ (p < .001)$			
Experiment 3					
Small P set	1/100	0.745 (.26)	1.332 (.88)	0.740 (.11)	1.182 (.39)
Medium P set	50/100	1.009 (.39)	1.337 (.84)	0.745 (.28)	0.909 (.61)
Large P set	90/100	1.248 (.39)	1.262 (.78)	0.737 (.41)	0.753 (.61)
		$\rho(N = 12) = .69\ (p < .025)$			

Note. $P(M_I)$ = .01. Figures in parentheses show the proportion of cards that were selected in Kirby's experiments. $SE(I_g)$ = scaled expected information gain; $P(M_I)$ = probability of the independence model.

with the *not-q* card are low, that is, 5/21 (this analysis assumes that each letter is equally probable and that there are 5 vowels and 21 consonants). Kirby suggested that this might be why subjects do not select the *not-q* card. Equation 9 predicts that these odds will increase if $P(p)$ is larger, and hence that subjects should choose the *not-q* card more frequently. Equation 9 predicts no changes for the *p, not-q*, and *q* cards, however.

In Kirby's (1994) Experiments 1 to 3, subjects checked whether a computer had made a mistake in generating cards with integers between 0 and 1,000 (or 0 and 100 in Experiments 2 and 3) on one side and either a "+" or "-" on the other side. In Experiment 1, subjects were told that the computer had an error rate of .01, and in Experiment 2, they were told it had an error rate of .1. In Experiment 1, the rules were *if there is a 0 on one side, there is a + on the other side* (small P set condition), and *if there is a number between 1 and 1,000 on one side, there is a + on the other side* (large P set condition). If each number is equally probable, then when 0 is the antecedent $P(p)$ is 1/1,001 and when any number between 1 and 1,000 is the antecedent $P(p)$ is 1,000/1,001. In his Experiments 2 and 3, Kirby used three values, so that $P(p)$ = 1/100, 50/100, or 90/100.

As he predicted, Kirby (1994) found that selections of the *not-q* card increased as $P(p)$ increased (see the figures in parentheses in Table 6). However, he also found unpredicted movements in the frequency of card selections for the other cards. As $P(p)$ increased, selections of the *p* card and *q* card decreased and selections of the *not-p* card increased, although the finding for the *q* card was not robust. Kirby considered a variety of possible

explanations for these effects that we now argue are direct consequences of our model of optimal data selection.

Model

From the perspective of optimal data selection, the failure of Kirby's (1994) model to predict the movements in card selections for the *p, not-p*, and *q* cards (indeed, the failure to predict that subjects should choose the *q* and *not-p* cards at all) is due to its exclusive focus on falsifying instances. In contrast, in our optimal data selection framework, a card can still be informative even though it could not yield a falsification. This permits us to model these data straightforwardly.

Our model predicts that the *not-q* card is informational to the extent that $P(p)$ is large. Our model therefore predicts Kirby's (1994) principle finding. It moreover predicts the other changes in the frequency of card selections that he observed for the *p, not-p*, and *q* cards. The independent variables Kirby manipulated in his Experiments 1 to 3 correspond closely to the parameters of our model. Kirby varied $P(p)$ directly. We assume that $P(q) = P(p)$. This assumption is reasonable because Kirby's materials are binary, that is, the antecedent is either 0 or 1 to 100 (or 1,000) and the consequent is "+" or "-." Staudenmayer and Bourne (1978) interpreted the effect of such material as leading to a biconditional interpretation, which is consistent with the assumption that $P(q) = P(p)$. Kirby provided specific information about the error rate; that is, in Experiment 1 it was .01 and in Experiment 2 it was .1. We assume that subjects use the error

rate as an estimate of the probability of the independence model. However, because an error rate of .1 is unreasonably high for a computer, we assume that subjects take $P(M_I)$ to be .01 for both Experiments 1 and 2. Using these parameter values we calculated $SE(I_g)$s for all cards in each condition of Kirby's Experiments 1 and 2. These appear in Table 6 together with the proportions of cards selected in Kirby's data (in parentheses). In his Experiment 3, Kirby provided no error rate information. Nonetheless, it is reasonable to assume that subjects regard computers as having a low error rate. Hence, we used the same parameter values to model Experiment 3 as in Experiment 2. As can be seen from Table 6, the fit between data and model is very good. The correlation for Experiment 1 is .87 and for Experiment 2 is .90. The fit in Experiment 3 is weaker (.69) but nonetheless significant. This may be due to subjects assuming a broader range of $P(M_I)$ values.

Therapy Experiments

Data

As their names suggest, these experiments (Wason, 1969; Wason & Golding, 1974; Wason & Johnson-Laird, 1970) involved therapeutic procedures to help subjects see where they were going "wrong" and to encourage them to adopt the falsificatory p, $not\text{-}q$ selection. The experimenter engaged subjects in a dialogue concerning their task performance, exposing them by degrees to inconsistencies "between their initial selections of cards and their subsequent independent evaluations of specific cards as falsifying or verifying the rule" (Wason & Johnson-Laird, 1972, p. 179). In Wason's study (1969), subjects performed an initial selection task and were then given three increasingly direct therapies: weak hypothetical contradiction, strong hypothetical contradiction, and concrete contradiction. Each therapy was followed by a further selection task to assess whether it was successful. Subjects also performed a final selection task, making five in all.

The therapies aimed to get agreement that a card with p and $not\text{-}q$ face falsified the rule. Three therapies laid an increasing emphasis on the $not\text{-}q$ card. The weak hypothetical contradiction therapy focused on the p card. The strong hypothetical contradiction therapy focused on the $not\text{-}q$ card. Hypothetical contradictions involved asking subjects what they thought could be on the hidden faces of the cards and getting them to agree that a p and $not\text{-}q$ instance falsified the rule. In the concrete contradiction therapy, the experimenter turned over the $not\text{-}q$ card to reveal a p on the hidden face.

These attempts at therapy were not wholly successful. By the final selection task only 42% adopted the falsificatory response. The steps by which these subjects moved to this response followed two main patterns (Wason, 1969), which later became the focus of various "insight" models (see below). In the first, subjects begin with an initial p card only response, they then move to a p, q, $not\text{-}q$ response and finally to a p, $not\text{-}q$ response (i.e., $p \Rightarrow p$, q, $not\text{-}q \Rightarrow p$, $not\text{-}q$). Notice that therapy on the $not\text{-}q$ card unexpectedly causes some subjects to choose the q card. In the second, subjects begin with an initial p and q response, they then move to a p, q, and $not\text{-}q$ response, and finally to a p,

$not\text{-}q$ response (i.e., p and $q \Rightarrow p$, q, $not\text{-}q \Rightarrow p$ and $not\text{-}q$; from Wason, 1969). Subjects successfully completed the second sequence less often (just 23.1% of subjects who made the initial p, q selection, rather than 62.5% of subjects who made an initial p selection [Wason, 1969]).

Various "insight" models attempted to explain these transition sequences (Bree, 1973; Bree & Coppens, 1976; Goodwin & Wason, 1972; Johnson-Laird & Wason, 1970a; Moshman, 1978; Smalley, 1974). A common feature of these models is that they postulate three levels of insight into the task, which subjects pass through sequentially and which correspond to the stages in the transition sequences identified in the last paragraph. Thus, in a state of "no insight," the subjects turn the p card and possibly also the q card; in a state of "partial insight," they turn the p, q, and $not\text{-}q$ cards; and in a state of "complete insight," they turn the p and $not\text{-}q$ cards (Johnson-Laird & Wason, 1970a).

Model

We explain these findings by the way subjects use the ordering in $SE(I_g)$ values (below) to determine which cards to select:

$$SE[I_g(p)] > SE[I_g(q)] > SE[I_g(not\text{-}q)] > SE[I_g(not\text{-}p)].$$

Let us consider the first transition sequence, which is the most striking, because therapy concerning the $not\text{-}q$ unexpectedly induces the selection of the q card. The subject initially chooses p only and assumes that the $SE(I_g)$ values of the other cards are not high enough to warrant their being turned. The effect of therapy is to persuade the subject that the experimenter considers that the $not\text{-}q$ is worth turning. If, as we have assumed, subjects consider card choice to be a monotonic function of $SE(I_g)$, then they should also turn the q card, because $SE[I_g(q)] > SE[I_g(not\text{-}q)]$. Hence, subjects subsequently choose p, q, and $not\text{-}q$. Subjects require further therapy to persuade them to violate monotonicity concerning the informativeness ordering and choose only p and $not\text{-}q$. We explain the second transition sequence in the same way, the only difference is that subjects' initial card selections are p, q. Thus, we can account for the main transition sequences observed in the data.

Subjects are reluctant to make the falsificatory response, even when strongly prompted to do so (only 42% of subjects finally make the p, $not\text{-}q$), because falsification requires them to violate the informativeness ordering. Our model does not directly predict that the first transition sequence should lead to more p, $not\text{-}q$ responses. However, it does suggest a possible explanation. Subjects' reluctance to move to the p, $not\text{-}q$ response stems from the tendency to want to turn the q card. It may, therefore, be difficult to persuade subjects not to turn the q card when they turned it initially. Hence, subjects who initially select p, q are less likely to complete the transition sequence than subjects who initially select p.

Reduced Array Selection Task (RAST)

Data

In a RAST, subjects choose between the q and $not\text{-}q$ options only (hence "reduced array"; Johnson-Laird & Wason, 1970b;

Wason & Green, 1984). The stimuli in the original RAST consisted of 30 colored shapes. The experimenter informs the subjects that there are 15 black shapes and 15 white shapes, each of which is a triangle or a circle. The shapes are in two boxes, one containing the white shapes and the other containing the black shapes. On being presented with a test sentence, for example, *All the triangles are black,* subjects have to assess the truth or falsity of the sentence by asking to see the minimum number of black or white shapes. In Johnson-Laird and Wason (1970b), although all subjects chose some confirmatory black shapes (no subject chose more than 9), they all chose all 15 potentially falsificatory white shapes. Thus, where subjects in effect perform multiple selection tasks, they tend to show falsificatory behavior.

Wason and Green (1984) reported a variant on this task. In one condition, the materials consist of cards colored on one half and depicting a shape on the other half. In this condition, the rule is disjoint, for example, *All the cards that have a triangle on one half are red on the other half* (the *All the triangles are red* rule they described as unified). In this condition, Wason and Green found that subjects predominantly select the *q* card. They also observed that their "experiments show relatively good performance in reasoning about conditional sentences using the RAST technique" (Wason & Green, 1984, p. 608). Even in the disjoint rule condition there was a falsificatory response rate of between 29% and 45%, compared with a falsificatory response rate as low as 4% in the standard selection task.

Model

The RAST makes explicit that the rule applies to a limited domain of cards or shapes that the experimenter describes as being in a box or a bag (or in Wason & Green [1984], "under the bar"). The experimenter also informs subjects that in this limited domain, there are equal numbers of *q* and *not-q* instances. It follows that $P(q) = .5$, violating the rarity assumption. At this value, $E[I_g(not\text{-}q)]$ is higher than $E[I_g(q)]$ (Figure 2), and hence our model predicts more *not-q* card selections than *q* card selections.

Our model does not directly predict that a disjoint rule reduces the facilitatory effect of the RAST (Wason & Green, 1984). It does suggest a possible explanation, however. The standard RAST rule specifies class inclusion, whereas the disjoint rule specifies a relationship between two distinct items. The latter suggests a causal relationship that cannot be restricted to the limited domain of cards "in the bag" or "in the box" (Goodman, 1954). Hence, the default rarity assumption may again be made in the disjoint condition.

Thematic Selection Task

Most recent work on the selection task has concentrated on how thematic content affects reasoning (e.g., Cheng & Holyoak, 1985, 1989; Cosmides, 1989; Evans, 1989; Gigerenzer & Hug, 1992; Girotto, Mazzocco, & Cherubini, 1992; Griggs & Cox, 1982; Jackson & Griggs, 1990; Johnson-Laird & Byrne, 1991, 1992; Manktelow & Over, 1987, 1990a, 1990b, 1991; Rumelhart, 1980). In the selection task, this work originated in the attempt to facilitate falsificatory reasoning (Johnson-Laird, Legrenzi, & Legrenzi, 1972; Wason & Shapiro, 1971). For example, subjects may have to imagine that they are an immigration official enforcing the rule that *If a passenger's form says "ENTERING" on one side, then the other side must include cholera* (Cheng & Holyoak, 1985; Cheng, Holyoak, Nisbett, & Oliver, 1986), or they may have to imagine that they are a tribal elder enforcing the rule that *If a man eats cassava root, then he must have a tattoo on his face* (Cosmides, 1989). Subjects are also given a rationale for enforcing the rule (the prevention of disease and that cassava root is a rare aphrodisiac that only married men, who have their faces tattooed, are allowed to eat). These thematic rules have typically facilitated the selection of the *p* and *not-q* cards.

Researchers now generally accept that these versions of the task address peoples' abilities at deontic reasoning, that is, reasoning concerning what ought to be the case (Manktelow & Over, 1987, 1990a, 1990b, 1991). In the abstract tasks, subjects "are asked to decide whether an indicative conditional is true or false, while in . . . [the deontic tasks] . . . they are asked whether a conditional obligation has or has not been violated. A conditional obligation, of course, is not falsified when it is violated" (Manktelow & Over, 1990b, p. 114). Thus, a subject's task in the deontic versions is very different to that confronted in the abstract versions of the selection task. However, we argue that the same probabilistic framework we used for the abstract selection task also applies to these data.

We describe these findings following Gigerenzer and Hug's (1992) classification by rule type and perspective. We also discuss recent work by Kirby (1994), who manipulated probabilities and utilities in a thematic task, in a separate section *Utilities and Probabilities of Fictional Outcomes.*

Rule Type and Perspective

Data. There are two dimensions on which the pattern of cards selected in the thematic selection task depends. The first is rule type. Cheng and Holyoak (1985) used rules like *If a passenger's form says "ENTERING" on one side, then the other side must include cholera,* which they described as "permissions." However, as Manktelow and Over (1987, 1990a, 1990b, 1991) observed, these rules are actually of the form of an obligation, that is, people who want to carry out the *action* described in the antecedent are obliged to satisfy the *condition* stipulated in the consequent. Obligations are of the form *if action* (*p*), *then must condition* (*q*). A corresponding permission would be *If a passenger's form includes cholera on one side, then they may enter the country.* Here people who have satisfied the condition described in the antecedent are permitted to perform the action described in the consequent. Permissions are of the form *if condition* (*p*), *then may action* (*q*). Notice that in going from an obligation to a permission, action and condition switch their clausal positions between antecedent and consequent of the conditional sentence.

The second dimension on which the pattern of card selections depends is the perspective a subject must adopt. Using an obligation rule, Cheng and Holyoak (1985) had subjects adopt the role of enforcers of the rule, that is, subjects had to imagine they

were immigration officials checking immigration forms. They found that subjects were more likely to select the p and not-q card combination under these conditions. Cosmides (1989) replicated this finding (as have Gigerenzer & Hug, 1992) and showed that from the enforcer's perspective, a permission (what Cosmides called a "switched social contract") led subjects to select the not-p and q cards. (Notice that both these responses still correspond to selection of the *action* and *not-condition* pair.) Using the obligation rule, Cosmides (1989) also asked subjects to adopt the role of inquirer's into whether a deontic rule was in force. She found similar results to those found in the abstract selection task.

Manktelow and Over (1991) were the first to argue that social role or perspective was an important factor in the deontic selection task. They induced response shifts between the p, not-q selections and the not-p, q selections by asking subjects to adopt different perspectives. They used a permission rule and two perspectives: what we have been calling the *enforcer's* perspective and what we shall refer to as the *actor's* perspective. For the enforcer's perspective Manktelow and Over (1991) found the same not-p and q card selections as Cosmides (1989; see also Gigerenzer & Hug, 1992), but for the actor's perspective they found that subjects predominantly chose the p and not-q cards. We illustrate the reason for this change using the permission form of the cholera rule. From the enforcer's perspective cheaters are people who try to enter the country without having been inoculated against cholera, that is, relative to the permission rule the not-p and q instances. However, for an actor, that is, someone trying to enter the country, the enforcer cheats if having had the inoculations the actor is still not let into the country, that is, again relative to the permission rule, the p and not-q instances. Gigerenzer and Hug (1992) showed that the actor's perspective on an obligation led subjects to select the not-p and q cards, because an obligation reverses the clausal position of action and condition. Gigerenzer and Hug (1992) also manipulated the same rules systematically along both dimensions (except the inquirer's perspective) for the first time.

Model. The thematic selection task requires that we refocus our existing probabilistic model away from *rule testing* and on to *rule use.* In modeling rule testing, we used our basic probability model defined by the dependence and independence matrices (Table 1) to calculate expected information gain. In modeling rule use, we use these probability models to calculate expected utilities and argue that subjects use the rules to maximize expected utility.

We assume that there is a small fixed cost for turning any card. This cost is implicit in the task, because the instructions say that subjects should only pick the cards that they "would have to" (Cheng & Holyoak, 1985) or "must" (Manktelow & Over, 1991) turn. We further assume that subjects associate particular utilities with particular card combinations, dependent on the perspective they adopt and on the particular rule.

The enforcer's goal is to discover instances of rule violation, that is, where the actor performs the action without satisfying the condition. We model this by assigning a positive utility to instances of rule violation that is larger than the cost of turning over a card (otherwise subjects would have no incentive to turn cards at all). Subjects in the enforcer's perspective associate no

Table 7

Utilities of Card Combinations for the Enforcer and Actor Perspectives

Perspective	Action	Not-action
Enforcer		
Condition	−0.1	−0.1
Not-condition	5 − 0.1	−0.1
Actor		
Condition	−0.1	5 − 0.1
Not-condition	−0.1	−0.1

other cards with positive utility. In particular, this means that whether someone performs the action when they satisfy the condition is not the enforcer's concern. The actor's goal is to discover instances of unfairness, where the enforcer disallows the action even though the actor satisfies the condition. We model this by assigning a positive utility to uncovering instances of unfairness that is larger than the cost of turning over a card. Subjects in the actor's perspective associate no other cards with positive utility. In particular, this means that whether someone performs the action when they do not satisfy the condition is not the actor's concern. The inquirer's goal is to discover whether the rule holds, just as in the abstract tasks. The inquirer has no direct involvement in the situation and therefore has no relevant utilities concerning it. We adopted the inquirer's perspective in modeling the abstract task, and the same analysis applies to the inquirer's perspective in thematic tasks.

We summarize the utilities assigned to the enforcer's and the actor's perspective in Table 7. In this table, we have adopted the convention that p corresponds to the condition and q to the action, that is, we assume a permission rule.[11]

We assume that subjects do not know whether the rule is being obeyed. For simplicity, we assign equal prior probabilities to the rule being obeyed and the rule being ignored, that is, $P(M_I)$ and $P(M_D)$ are equal. The interpretation of the parameter $P(M_I)$ changes between the abstract and thematic tasks. In the abstract task, it represented the probability that the independence model is true of the world. In the thematic task, it represents the probability that an individual is disobeying the rule. With this modification, we model the task in exactly the same way as in the abstract selection task, except that we introduce utilities with respect to the model.

To model the card selections in the enforcer's and actor's perspective, we calculate expected utilities for each card as follows (we use the abbreviations "act" for action and "con" for condition):

$$EU(con) = P(act|con)U(con, act) + P(\overline{act}|con)U(con, \overline{act}) \quad (10)$$

$$EU(\overline{con}) = P(act|\overline{con})U(\overline{con}, act) + P(\overline{act}|\overline{con})U(\overline{con}, \overline{act}) \quad (11)$$

$$EU(act) = P(con|act)U(con, act) + P(\overline{con}|act)U(\overline{con}, act) \quad (12)$$

$$EU(\overline{act}) = P(con|\overline{act})U(con, \overline{act}) + P(\overline{con}|\overline{act})U(\overline{con}, \overline{act}). \quad (13)$$

---

[11] We could have equally adopted an obligation rule and hence let p be the action and q the condition. However, this is only a convention, and the expected utilities come out the same either way.

Table 8

Mean Expected Utilities for Each Card Face (Action, Not-Action, Condition, and Not-Condition) for the Enforcer's Perspective and the Actor's Perspective

Card face	Enforcer	Actor
Action	1.20	−0.10
Not-action	−0.10	2.31
Condition	−0.10	2.23
Not-condition	1.03	−0.10

Where the conditional probabilities $P(x|y)$ are the expected values calculated with respect to the two models:

$$P(x|y) = P(x|y, M_I)P(M_I) + P(x|y, M_D)P(M_D). \quad (14)$$

In Equations 10 to 13, the expected utility of each card is calculated as the weighted sum of the probabilities of each possible outcome given the visible face of the card. The weights are the utilities, $U()$, of each outcome.

We derived expected utilities for each card by sampling points in the parameter space defined by $P(p)$ and $P(q)$ at intervals of .1 in the range .1 to .9, with the utilities specified above (−.1 fixed cost and 5 for the target). We sampled over a whole range of values for $P(p)$ and $P(q)$ because for deontic rules it is not reasonable to prejudge rarity. For example, in monitoring passengers, whether most or only some of the passengers are entering will depend on factors such as the particular flight. If Manila is the flight destination, then most passengers will be entering. However, if the passengers are on the long haul flight from London to Sydney, then only some passengers may be entering, the rest will be in transit. We show the mean expected utilities in Table 8 for the enforcer's perspective and for the actor's perspective.

The enforcer seeks the case where the actor performs the action but does not satisfy the condition. In the model, selecting the face that denotes the action being performed and the face that denotes the condition not being satisfied maximizes expected utility. Table 8 shows that only the cards showing the *action* and *not-condition* face have positive expected utilities. Hence, subjects should turn only these cards. For an obligation, *if p* (action), *then must q* (condition), this corresponds to selecting the *p* and *not-q* cards. For a permission, *if p* (condition), *then may q* (action), this corresponds to selecting the *not-p* and *q* cards.

The actor seeks the case where, although the actor satisfies the condition, the enforcer disallows the action. In the model, selecting the face that denotes the condition being satisfied and the face that denotes the action not being taken maximizes expected utility. Table 8 shows that only the cards showing the *not-action* and *condition* face have positive expected utilities. Hence, subjects should turn only these cards. For an obligation, *if p* (action), *then must q* (condition), this corresponds to selecting *not-p* and *q* cards. For a permission, *if p* (condition), *then may q* (action), this corresponds to selecting the *p* and *not-q* cards.

In summary, our model makes the predictions for card selections in the deontic selection task shown in Table 9. These pre-

dictions agree perfectly with the results of the studies indicated. We also predict that a permission rule with an inquirer's perspective will lead to the standard abstract results, because from the inquirer's perspective our standard abstract model should apply.

Utilities and Probabilities of Fictional Outcomes

Data. Kirby (1994) has recently demonstrated that the utilities and probabilities of outcomes affect card selections in the deontic selection task. Equation 9 again forms the basis of his analysis. In his Experiment 4, Kirby used a drinking age deontic rule as used by Griggs and Cox (1982): *if a person is drinking beer, then the person must be over 21 years of age.* Kirby used the following cards, "drinking beer" (*p*), "drinking ginger ale" (*not-p*), "22 years of age" (*q*), and "19 years of age" (*not-q*). This rule is an obligation rule, and subjects must adopt the enforcer's perspective that predicts the *p* and *not-q* response. Kirby argued that the high frequency of *not-q* card selections found by Griggs and Cox (1982) may be due to the high probability of finding a 19-year-old drinking beer. He therefore provided two additional *not-q* cards that varied this probability, "12 years of age" and "4 years of age"—the younger the person, the less likely they are to be drinking beer.

In the same experiment, Kirby (1994) varied the utilities of making correct and incorrect decisions. In a DON'T CHECK condition, the instructions read, "However, keep in mind that your employer does not want you to offend innocent customers, and you could be fired if you check an innocent person." From Equation 9 these instructions should increase the cost of a FA. Kirby therefore predicted an overall decrease in the number of cards selected. In a DON'T MISS condition, the instructions read, "However, keep in mind that your employer is very concerned about illegal drinking, and you could be fired if you miss a guilty person." From Equation 9 these instructions should increase the cost of a miss. Kirby therefore predicted an overall increase in the number of cards selected. In a CHECK condition, the instructions read, "However, keep in mind that your employer is very concerned about illegal drinking, and you could receive a large bonus if you catch a guilty person." From Equation 9 these instructions should increase the benefit of a hit. Kirby therefore predicted an overall increase in the number of cards selected. He compared these data to a BASELINE condition with no manipulation of these utilities.

In his Experiment 4, consistent with prediction, Kirby (1994) observed a trend for fewer selections for *not-q* cards with a lower probability of an inconsistent outcome. Moreover, the DON'T CHECK condition led to fewer card selections than the BASELINE condition, and the DON'T MISS condition led to more card selections than the BASELINE condition, as predicted. Similar effects were not observed for the CHECK condition. Kirby argued that this was because this condition involved a less extreme benefit, and subjects weight costs more than benefits (Kahneman & Tversky, 1979). Kirby's results for the BASELINE, the DON'T CHECK, and the DON'T MISS conditions appear in parentheses in Table 10.

Model. Modeling Kirby's (1994) data is straightforward. First, in the abstract task, we set the parameter $P(M_I)$ to the

Table 9

Patterns of Card Selections Observed in the Thematic Selection Task for Different Rule Types, and Perspectives Indicating the Studies Reporting These Results

	Perspective		
Rule type	Enforcer	Actor	Inquirer
Obligation	p, not-q (Cheng & Holyoak, 1985; Cosmides, 1989; Gigerenzer & Hug, 1992)	not-p, q (Gigerenzer & Hug, 1992)	standard abstract result: $p > q > not\text{-}q > not\text{-}p$ (Cosmides, 1989)
Permission	not-p, q (Cosmides, 1989; Manktelow & Over, 1991; Gigerenzer & Hug, 1992)	p, not-q (Manktelow & Over, 1991; Gigerenzer & Hug, 1992)	standard abstract result: $p > q > not\text{-}q > not\text{-}p$ (prediction)

error rate in Kirby's Experiments 1 to 3. For the thematic task, this parameter reflects the probability that an individual is disobeying the rule. We therefore varied this parameter to model the effect of the various ages of potential violators (*not-q* cards). We set $P(M_I)$ to .4 for the 4-year-olds and then incremented by .1 for the 12-year-olds ($P[M_I] = .5$) and the 19-year-olds ($P[M_I] = .6$). These values seemed reasonable because even though 4-year-olds in general are unlikely to be drinking beer, the probability of 4-year-olds in a bar drinking is far higher. Certainly the subjects in Kirby's Experiment 4 felt it necessary to check the 4-year-olds, the proportion of these cards being turned never dropping below .39.

To capture the effects of the different instructions we varied the utilities specified in Table 7 for the enforcer's perspective. The DON'T CHECK condition increases the cost of a false alarm. We model this directly by increasing the costs for all cells

other than the *action, not-condition* cell, because any of the outcomes corresponding to these cells represents a false alarm. We cannot increase costs too much, however, otherwise they will outweigh the benefits for all possibilities and enforcers will carry out no checks at all. We therefore increased the costs from $-.1$ to $-.5$. We doubt whether subjects make the distinction between a cost for a miss and a benefit for a hit. It seems more reasonable to assume that the cognitive interpretation of costs for misses is benefits (failure to incur a cost) for hits. Therefore, the DON'T MISS condition is a more extreme version of the CHECK condition. We, therefore, model the DON'T MISS condition by increasing the utility of the *action, not-condition* cell in Table 7 for the enforcer's perspective.

We illustrate the behavior of the model with $P(p) = P(q) = .5$. (Any pair of values in the range .1 to .9 displays the same behavior in response to variations in $P(M_I)$ and the utilities.) Table 10

Table 10

Expected Utilities for Each Card With P(p) = P(q) = .5 in Each Condition of Kirby's Experiment 4

		Card				
Condition	$P(M_I)$	Beer (p)	Ginger ale ($not\text{-}p$)	22 years old (q)	($not\text{-}q$)	M
BASELINE[a]						
19 years	.6	1.78 (.91)	−.1 (0)	−.1 (.07)	1.40 (.86)	
12 years	.5	1.57	−.1	−.1	1.15 (.76)	
4 years	.4	1.33	−.1	−.1	0.90 (.71)	(.55)
DON'T CHECK[b]						
19 years	.6	1.38 (.80)	−.5 (.05)	−.5 (.08)	1.00 (.81)	
12 years	.5	1.17	−.5	−.5	0.75 (.47)	
4 years	.4	0.93	−.5	−.5	0.50 (.39)	(.44)
DON'T MISS[c]						
19 years	.6	2.53 (.92)	−.1 (.05)	−.1 (.03)	2.00 (.97)	
12 years	.5	2.23	−.1	−.1	1.65 (.86)	
4 years	.4	1.90	−.1	−.1	1.30 (.80)	(.60)

Note. $\rho(N = 18) = .94$ ($p < .0001$). Numbers in parentheses indicate the proportion of cards selected in Kirby's experiment.
[a] Cost = −.1; benefit = 5.   [b] Cost = −.5; benefit = 5.   [c] Cost = −.1; benefit = 7.

shows the expected utilities for each card for the BASELINE, the DON'T CHECK, and the DON'T MISS conditions. The fit between data and model was good with a correlation of .94 ($p <$ .0001). As in the abstract task, our model captures effects that Kirby's (1994) model cannot explain. From Equation 9, Kirby must predict increases and decreases in card selections for all cards. However, although there were overall increases and decreases according to Equation 9 in Kirby's data (see Table 10, right-hand column), these effects were only in line with prediction for the p and not-q cards. There were no significant changes in the proportion of cards selected for the not-p or q cards, and where Kirby found differences they were counter to the predictions of Equation 9. In contrast, our model predicts no changes for these cards in response to Kirby's manipulations, which is consistent with Kirby's results.

Summary for Thematic Tasks

Our account of the thematic selection task is consistent with some recent proposals. It supplements our standard probability model with a maximum expected utility account of the role of perspectives. Following Manktelow and Over (1991, 1992) and Gigerenzer and Hug (1992) the notion of a perspective is the main explanatory concept. Moreover, we have built on Manktelow and Over's (1991) qualitative explanation of the influence of perspectives in terms of subjective utilities. Our emphasis on the distinction between rule testing and rule use is also consistent with Jackson and Griggs' (1990) finding that the checking (rule use) context is a more important factor in reasoning in these tasks than their deontic nature. Our proposals are also in the spirit of Gigerenzer and Hug's (1992) expectation that "the two fields [i.e., deductive and probabilistic reasoning] will converge in the next few years" (p. 169). In this light, perhaps the most important feature of our model is how well it accounts for Kirby's (1994) recent data, where both probabilities and utilities are explicitly varied in a deontic task.

Discussion

We begin the discussion by informally summarizing our model and showing that it provides a rational analysis of the selection task. We then compare our model to other accounts of the selection task and to other accounts of reasoning more generally. Finally, we relate our work to the wider program of understanding human behavior as rational.

Summary

We view the standard abstract selection task as an inductive reasoning task. Subjects must choose which card-turning experiments they expect to yield the most information about which of two hypotheses are true. One hypothesis is that a dependency of the form *if p, then q* exists, the other is a foil hypothesis, that p and q are independent. We defined expected information gain as the expected decrease in information-theoretic uncertainty between these two hypotheses in response to some data. We formalized expected information gain, $E(I_g)$, using the theory of optimal data selection from statistics. We then assumed that

card selection frequencies are a monotonic function of the expected information gain associated with each card.

The model has only three free parameters: the prior probability of the independence model, $P(M_I)$; the probability of p, $P(p)$; and the probability of q, $P(q)$. We explain the majority of the effects on the abstract selection task by assuming that p and q are rare by default and that experimental manipulations influence these parameters by moving them away from their default values. When parameters $P(p)$ and $P(q)$ are low, the ordering in expected information gain corresponds to the standard pattern of card selection frequencies ($p > q > not$-$q > not$-p). With the standard results as a baseline, we explained the associations between card selections by making two minimal assumptions about the nature of the decision process that translates expected information gains into card selections. These assumptions allowed us to compute scaled $E(I_g)$s that showed the same pattern of associations found in the empirical data. We accounted for the negations paradigm selection task, using the same assumptions and by allowing $P(p)$ and $P(q)$ to vary according to an account of contrast sets. Our account also captures all of Kirby's (1994) recent data where he explicitly varies $P(p)$. Furthermore, our model accounts for unexpected selection transitions in the therapy experiments and for the facilitation of the logical response in the RAST.

Following Manktelow and Over (1987, 1990a, 1990b, 1991), we assume that the strikingly different results observed in many thematic selection tasks stem from the deontic nature of the rules used. Such rules are not hypotheses to be tested but rules that must be obeyed. We model the thematic task using decision theory together with the same basic probability models used to model the abstract task. Subjects have utilities concerning various outcomes that depend on their perspective toward the rule. We assume that they choose cards to maximize expected utility. The assumption that subjects follow this rational policy captures a broad range of data on the thematic selection task. In particular, it captures the effects of perspective and rule type that have been much studied recently; it also captures Kirby's (1994) data, where utilities and probabilities have been explicitly varied in a deontic selection task.

Relation to Rational Analysis

Our model provides a rational analysis of the selection task in the sense of Anderson (1990, 1991a, 1991b; Anderson & Milson, 1989). According to Anderson (1990), rational analysis involves six steps:

1. Specify precisely the goals of the cognitive system.
2. Develop a formal model of the environment to which the system is adapted.
3. Make minimal assumptions about computational limitations.
4. Derive the optimal behavior function given 1–3 above.
5. Examine the empirical evidence to see whether the predictions of the behavioral function are confirmed.
6. Repeat, iteratively refining the theory.

We assume that the goals (1) involve selecting relevant data, that the structure of the environment (2) is given by the frequency of properties occurring in that environment, and that

the costs (3) are incurred in looking at irrelevant data. We derived an optimal behavioral function using Bayesian optimal data selection (4) and compared this with the empirical evidence (5). In summary, our model demonstrates the utility of Anderson's (1990) approach by showing how it can organize data on human reasoning that has previously seemed the most recalcitrant to rational explanation.

Relations to Theories of Deductive Inference

We deal with the relation of our model to theories of deductive inference in two parts. First, we look at the relations between our model and theories of deductive inference taken to account for the same data. Next we look at some recent probabilistic approaches in reasoning research to which our account is more closely related.

Theories of deductive inference. Evans (1991) has proposed a taxonomy of deductive reasoning theories that divides the four principle approaches into two classes: those that deal with the issue of deductive competence and those that do not. On the one hand, mental logic accounts (Braine, 1978; Rips, 1983, 1990) and mental models (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991) are theories of deductive competence. On the other hand, domain-specific approaches, such as pragmatic reasoning schemas (Cheng & Holyoak, 1985) and social contract theory (Cosmides, 1989), and heuristic or "relevance" approaches (Evans, 1983, 1984, 1989), account for the content effects and biases found in deductive reasoning tasks. We look first at the relations between our model and accounts of deductive competence.

Accounts of deductive competence. According to the mental logic approach (e.g., Braine, 1978; Rips, 1983, 1990), deductive reasoning involves an abstract logical competence implemented in something like the syntactic rules of a standard logical system. This approach does not attempt to explain the selection task. Rips (1990) argued that the selection task is a "loose" probabilistic task, rather than a "tight" deductive task. We agree and note that our theoretical account shows how the selection task can be modeled in probabilistic terms. Consequently, selection task data cannot be used as an argument against a mental logic, and hence the case against a mental logic is weaker than perhaps it once was.

The mental models (e.g., Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991) framework proposes that people do not reason by manipulating syntactic rules but by manipulating representations of the semantic contents of an argument. Mental models has problems with accounting for the data on the selection task. For example, Johnson-Laird and Byrne (1991) claimed that subjects who do not turn the *not-q* card do not represent it in their mental model. This suggests that when an experimental manipulation draws explicit attention toward this card, subjects should select it. However, in the therapy experiments, where the experimenter focuses attention on the *not-q* card and its falsifying properties, the majority of subjects still do not select it.

Problems of predictive failure to one side, mental models has most difficulty in accounting for the influence of probabilities and utilities on reasoning performance. Johnson-Laird and

Byrne (1992) argued that such factors only enter into the construction of appropriate mental models, and hence they need not incorporate these factors in their framework. Garnham (1993) has attempted a similar argument in defending mental models theory from the criticism that it fails to account for everyday inference (Chater, 1993; Chater & Oaksford, 1993; Oaksford, 1993; Oaksford & Chater, 1991, 1992, 1993; Oaksford, Chater, & Stenning, 1990). In our view, theories of everyday inference (Galotti, 1989) will involve how factors such as probabilities and utilities interact with reasoning processes (see also Gigerenzer & Hug, 1992). However, mental models theorists appear to believe that they already have a theory of everyday inference (Garnham, 1993; Johnson-Laird & Byrne, 1991). This is only true if the mechanisms that construct just the right kinds of mental model are assumed as primitive. As Chater and Oaksford (1993; see also Kirby, 1994) observed, "this line of reasoning has, in Russell's phrase, all the virtues of theft over honest toil" (p. 81)—most interesting inferential processes are presupposed and not explained (Oaksford & Chater, 1993).

In closing this section, we observe that, in contrast to Kirby's (1994) account, our model does not require an abstract logical competence. Kirby's signal-detection theory analysis defines hits as revealing a logically inconsistent outcome and hence presupposes an understanding of conditional logic. In our model, we characterize the hypotheses probabilistically by the matrices in Table 1. Thus, our account divorces selection task performance more completely from theories of deductive competence than Kirby's account.

Accounts of biases and content effects. Our principle criticism of approaches put forward to account for biases and content effects is that they lack the generality of our model. Domain-specific approaches such as pragmatic reasoning schema theory (Cheng & Holyoak, 1985) and social contract theory (Cosmides, 1989) deal only with the data from the thematic task. Both these accounts assume domain-specific rules for checking for violators in deontic contexts. The main difference is that on pragmatic reasoning schema theory these rules are learned, whereas on social contract theory they are innate. The emphasis on domain-specific information is compatible with our account (see also Kirby, 1994). Domain-specific knowledge may influence the parameters in our model, and the utilities subjects use, as noted in the Thematic Selection Task section.

The heuristic approach of Evans (1983, 1984, 1989) deals only with the data from the abstract task. Evans (1983, 1984, 1989) proposed that various heuristic processes involved in language understanding may explain the biases observed in the abstract selection task. In particular, Evans has applied this approach to matching bias, which we discussed in the Negations Paradigm Selection Task section. However, Oaksford and Stenning (1992) have shown that the particular heuristics Evans proposes are unnecessary to account for these data.

Relation to Probabilistic Approaches

There have been some loose probabilistic approaches (Rips, 1990) to the selection task (Fischhoff & Beyth-Marom, 1983; Klayman & Ha, 1987; Rips, 1990; and in the related Wason [1960] 2-4-6 task [Hoenkamp, 1989]). Fischhoff and Beyth-

Marom (1983) and Rips (1990) both adopted a Bayesian approach, but as part of more general frameworks for looking at hypothesis testing and loose reasoning tasks, respectively. In consequence, neither of these Bayesian approaches went beyond accounting for the standard p and q card selections, which is perhaps why they have had little influence on selection task research. Similar comments apply to Klayman and Ha (1987, 1989) who generalized their "positive test strategy" from Wason's 2-4-6 task to the selection task. They based this strategy on a demonstration that the likelihood of finding disconfirming evidence was higher when using positive instances of a hypothesis than when using negative instances of a hypothesis (as in falsification). In summary, these earlier accounts pointed the way to our probabilistic treatment of the selection task but were never themselves sufficiently developed to account for the range of data reviewed here.

Only Kirby (1994) tried to explain both the abstract and the thematic tasks from a probabilistic or decision theoretic perspective. The main failing of Kirby's otherwise excellent work is that his theory (Equation 9) can explain so little of his important data. This is because his analysis requires that subjects concentrate solely on finding falsifying instances. In contrast, our Bayesian analysis explains all of Kirby's data straightforwardly.

Interestingly, our model bears close relations to probabilistic approaches to causal reasoning, an area that until now has been treated as unrelated to the selection task. Anderson (1990) also used Bayesian model comparison in his model of causal estimation that provides a rational explanation of biases in the analysis of 2 × 2 contingency tables (e.g., Schustack & Sternberg, 1981). Cheng and Novick (1990, 1992) have also taken a probabilistic approach to both causal inference and to causal attribution (see, e.g., McArthur, 1972) in their probabilistic contrast model. Cheng and Novick (1990, 1992) did not propose a full Bayesian treatment of these data. However, their emphasis on probabilistic contrasts is similar to our emphasis on information gain in deciding between hypotheses. In both cases, subjects are assumed to concentrate on probabilistic differences. We may, moreover, be able to derive the probabilistic contrast model from our Bayesian framework. We can compare an independence model and a model (or family of models) in which a contingency holds (or parameterized family of models, each representing a different contingency reliability) not just for a single data point (as in the current analysis) but for an entire set of data. This could provide the basis for a normative analysis of experiments on contingency judgments, causal reasoning, and causal attribution. This opens up the exciting possibility of unified rational explanations of formerly disparate phenomena in the reasoning field.

Finally, our model is consistent with a growing trend in accounting for putative biases in inferential behavior using rational probabilistic models. Anderson's work is the most comprehensive such approach, applying Bayesian methods to a variety of cognitive phenomena (Anderson, 1990, 1991a, 1991b; Anderson & Milson, 1989). Gigerenzer (Gigerenzer, Hell, & Blank, 1988; Gigerenzer, Hoffrage, & Kleinbölting, 1991; Gigerenzer & Murray, 1987) has also applied probability theory to explaining biases in reasoning tasks, and Cheng and Novick's (1990, 1992) work is also consistent with this trend.

Rationality

There are two issues about rationality that require discussion. The first concerns the actual parameter values that we have chosen in our analyses and whether we can provide them with a rational justification. The second concerns the normative status of our rational analysis.

Parameter values. We have explained the data on the selection task by assuming that p and q are rare by default and that experimental manipulations influence the parameters $P(p)$ and $P(q)$ by moving them away from their default values. An initial and important point is that our model organizes a wide range of data in a single theoretical framework. This argues strongly that subjects behave as Bayesians with the rarity assumption. This in itself is an important discovery, even if we could not rationally justify the rarity assumption. Testing the validity of this assumption will require an environmental analysis of the type Anderson (1990) proposed. However, we argue that there is evidence to support the view that most lexicalized properties refer to objects and events that are rare in our environment. In consequence, subject's behavior in the selection task may be optimally adapted to that environment and hence rational.

First, note that no other parameter values are better justified. For example, the principle of indifference (Keynes, 1921) that $P(q) = P(p) = .5$ is only reasonable on the assumption of complete ignorance. However, subjects have extensive prior experience with other conditional rules. If these generally relate properties that respect rarity, then it is reasonable for subjects to extrapolate from prior experience and assume that a novel task rule also respects rarity. Other possibilities are equally questionable. For example, Kirby (1994) argued that the probability of finding a vowel (p) on the back of an odd number (*not-q*) is low because there are 5 vowels but 21 consonants. However, the level of letter *types* may not be the relevant level at which to assess these probabilities. It could equally be the level of letter *tokens* in experience that is the determining factor.

The rarity assumption organizes data from more than the selection task. We mentioned above that in the 2-4-6 task, Klayman and Ha (1987) showed that positive tests were more likely to yield falsifying evidence than negative tests. This result also relies on a rarity assumption, or what Klayman and Ha called a *minority phenomenon* assumption. That is, the properties that figure in hypotheses about their causes are in the minority. For example, AIDS has only an incidence of about 10^{-4} to 10^{-5} in the population. A hypothesis of the form *if you have contracted HIV, then you will develop AIDS* will therefore respect the rarity assumption. This is because scientists are unlikely to put much stock in this hypothesis if $P(\text{HIV}) > P(\text{AIDS})$ (this is a further application of our constraint that $P[q] \geq P[p]$). Furthermore, Anderson's (1990) work on causal inference indicates that subjects make a rarity assumption in causal estimation from 2 × 2 contingency tables. In fitting parameters to Schustack and Sternberg's (1981) data, Anderson derived expected prior probabilities of an effect of .27 and of a cause of .25.[12] Given a causal (*if cause, then effect*) relation, these results confirm the rarity assumption and the reasonable constraint that although $P(q)$ is greater than $P(p)$, it can only be marginally greater. In summary,

the rarity assumption appears capable of organizing a great deal of data on human reasoning.

Normative status. As we mentioned in the introduction, Anderson (1990) drew the distinction between normative and adaptive rationality (see also Evans, 1993; Stich, 1990). Normative rationality concerns reasoning according to the rules of a formal logico-mathematical theory. Following such rules provides the standard account of rationality going back to Plato (see Brown, 1989). Adaptive rationality concerns whether behavior is optimally adapted to the environment. We have shown that in the selection task, subjects' behavior can be regarded as rational in the second sense, that is, as optimally adapted to an environment where the properties that enter into predictive relations are rare. Although we have used a normative mathematical theory to derive this rational analysis, there is no requirement that people achieve this optimal adaptation by following the rules of the normative theory. Hence, although our account argues for the adaptive rationality of reasoning on the selection task, it need not address the question of normative rationality.

However, although a rational analysis does not require that people make Bayesian calculations, it does not preclude it either. A range of views is possible. At one extreme, we can view the calculations involved in deriving the rational analysis as specifying a set of mental operations carried out by the subject. This view attributes people with sophisticated, though not necessarily explicit (Reber, 1989), probabilistic reasoning abilities. It also corresponds to the view that people are not only adaptively rational but are also normatively rational. At the opposite extreme, as we have mentioned, a rational analysis may just specify which behavior is optimal and remain neutral about the mental operations underlying that behavior. The reason that people conform to our analysis of the selection task might be due to innate constraints or learning, rather than sophisticated probabilistic calculation. Between these two extremes, that all the calculations of the rational analysis are internally computed and that none are, lies a smooth continuum of intermediate positions, which assumes that some aspects of the analysis are calculated internally and others are not.

The view taken toward rational analysis has behavioral and computational significance. Insofar as people calculate optimal behavior internally, a subject's knowledge of the specifics of the task can influence those calculations. For example, in the RAST, we assume that the way in which the materials violate the rarity assumption influences a subject's behavior. If subjects performed no calculation, but simply applied learned or innate strategies, then it is unlikely that such parameter changes would affect their performance. It is, of course, possible that subjects choose between various strategies that do not involve calculation, depending on the specifics of the situation. Nonetheless, in general, the more flexible a subject's behavior to relevant aspects of the task, the stronger the case for internal calculation, and the less plausible noncomputational strategies. We will need to conduct further empirical work to assess which aspects of our rational analysis of the selection task people internally calculate and which they have prestored.

Insofar as people make internal calculations of our rational analysis, we must consider the computational feasibility of those

calculations. The calculations of our analysis of the selection task are very simple. However, as we have argued elsewhere (Chater & Oaksford, 1990; Oaksford & Chater, 1991, 1992, 1993), plausible reasoning theories must "scale up" from laboratory reasoning tasks to everyday inferential processes. Simple Bayesian calculations rapidly lead to the notorious combinatorial explosion (e.g., Charniak & McDermott, 1985). Recently Pearl (1988) has proposed a novel and more tractable implementation of Bayesian inference using Bayesian networks. However, this method too does not scale well (Dagum & Luby, 1993). These problems are not specific to probabilistic inference, but apply equally to logical reasoning (Oaksford & Chater, 1991).

The problems of computational tractability suggest that a scaled-up rational analysis would have to pay considerably more attention to computational limitations (Step 3 in Anderson's [1990] account of rational analysis) than is required for modeling laboratory tasks. In Simon's (1959) terms, this means that people should be modeled as having bounded rationality (see also Oaksford & Chater, 1992, 1993).

Conclusions

We have shown how to develop a rational analysis of the selection task, which accords closely with empirical data, using a Bayesian account of hypothesis testing. This account contrasts sharply with the standard falsificationist model. The poor fit between this model and the empirical data has led to doubts about whether humans are rational. We suggest that people are rational but that we must define rationality in terms of optimal performance in real-world, uncertain, inductive tasks, rather than purely in terms of deductive logic. Clarifying the detailed relationship between normative theory and observed behavior suggests a program of empirical investigation and theoretical generalization to related tasks, the success of which remains to be decided. Nonetheless, our model establishes that subjects' behavior while performing the selection task need have no negative implications for human rationality.

---

[12] Anderson (1990, p. 160) reported expected prior probabilities of an event (E), $P(E)$, of .27; of an event given the presence of a cause (C), $P(E|C)$, of .71; and of an event in the absence of a cause, $P(E|not-C)$, of .12. By elementary probability theory, $P(C)$ therefore equals .25.

References

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

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., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96,* 703–719.

Barwise, J., & Perry, J. (1983). *Situations and attitudes.* Cambridge, MA: MIT Press.

Bayes, T. (1958). An essay towards solving a problem in the doctrine of chances. *Biometrika, 45,* 293–315. (Reprinted from *Philosophical Transactions of the Royal Society, 53,* 370–418)

Beattie, J., & Baron, J. (1988). Confirmation and matching biases in

hypothesis testing. *Quarterly Journal of Experimental Psychology, 40A*, 269-297.

Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85*, 1-21.

Bree, D. S. (1973). The interpretation of implication. In A. Elithorn & D. Jones (Eds.), *Artificial and human thinking*. New York: Elsevier Science.

Bree, D. S., & Coppens, G. (1976). The difficulty of an implication task. *British Journal of Psychology, 67*, 579-586.

Brown, H. I. (1989). *Rationality*. London: Routledge.

Charniak, E., & McDermott, D. (1985). *An introduction to artificial intelligence*. Reading, MA: Addison-Wesley.

Chater, N. (1993). Mental models and non-monotonic reasoning. *Behavioral and Brain Sciences, 16*, 340-341.

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 and Language, 8*, 72-89.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391-416.

Cheng, P. W., & Holyoak, K. J. (1989). On the natural selection of reasoning theories. *Cognition, 33*, 285-313.

Cheng. P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic vs. syntactic approaches to training deductive reasoning. *Cognitive Psychology, 18*, 293-328.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58*, 545-567.

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review, 99*, 365-382.

Cherniak, C. (1986). *Minimal rationality*. Cambridge, MA: MIT Press.

Chrostowski, J. J., & Griggs, R. A. (1985). The effects of problem content, instructions and verbalization procedure on Wason's selection task. *Current Psychological Research and Reviews, 4*, 99-107.

Churchland, P. S. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences, 4*, 317-370.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*, 187-276.

Dagum, P., & Luby, M. (1993). Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence, 60*, 141-153.

Davidson, D. (1984). On the very idea of a conceptual scheme. In D. Davidson (Ed.), *Inquiries into truth and interpretation* (pp. 183-198). London: Oxford University Press.

Dennett, D. C. (1987). Making sense of ourselves. In D. C. Dennett (Ed.), *The intentional stance* (pp. 83-102). Cambridge, MA: MIT Press.

Duhem, P. (1954). *The aim and structure of physical theory*. Princeton, NJ: Princeton University Press. (Original work published 1914)

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

Evans, J. St. B. T. (1977). Toward a statistical theory of reasoning. *Quarterly Journal of Experimental Psychology, 29*, 621-635.

Evans, J. St. B. T. (1982). *The psychology of deductive reasoning*. London: Routledge.

Evans, J. St. B. T. (1983). Linguistic determinants of bias in conditional reasoning. *Quarterly Journal of Experimental Psychology, 35A*, 635-644.

Evans, J. St. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75*, 451-468.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hillsdale, NJ: Erlbaum.

Evans, J. St. B. T. (1991). Theories of human reasoning: The fragmented state of the art. *Theory and Psychology, 1*, 83-106.

Evans, J. St. B. T. (1993). Bias and rationality. In K. I. Manktelow & D. E. Over (Eds.), *Rationality* (pp. 6-30). London: Routledge.

Evans, J. St. B. T., & Lynch, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology, 64*, 391-397.

Federov, V. V. (1972). *Theory of optimal experiments*. San Diego, CA: Academic Press.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review, 90*, 239-260.

Fodor, J. A., Fodor, J. D., & Garrett, M. F. (1975). The psychological unreality of semantic representations. *Linguistic Inquiry, 6*, 515-531.

Galotti, K. M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin, 105*, 331-351.

Garnham, A. (1993). Is logicist cognitive science possible? *Mind and Language, 8*, 49-71.

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

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43*, 127-171.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.

Girotto, V., Mazzocco, A., & Cherubini, P. (1992). Judgments of deontic relevance in reasoning: A reply to Jackson and Griggs. *Quarterly Journal of Experimental Psychology, 45A*, 547-574.

Glass, G., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

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

Goodman, N. (1954). *Fact, fiction and forecast*. Cambridge, MA: Harvard University Press.

Goodwin, R. Q., & Wason P. C. (1972). Degrees of insight. *British Journal of Psychology, 63*, 205-212.

Griggs, R. A. (1984). Memory cueing and instructional effects on Wason's selection task. *Current Psychological Research and Reviews, 3*, 3-10.

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology, 73*, 407-420.

Griggs, R. A., & Cox, J. R. (1983). The effects of problem content and negation on Wason's selection task. *Quarterly Journal of Experimental Psychology, 35A*, 519-533.

Hampton, J. A. (1989, June). *Negating noun concepts*. Paper presented at the Edinburgh Round-Table on the Mental Lexicon, University of Edinburgh, Edinburgh, Scotland.

Hill, B. M., & Hunter, W. G. (1969). A note on designs for model discrimination: Variance unknown case. *Technometrics, 11*, 396-400.

Hoch, S. J., & Tschirgi, J. E. (1985). Logical knowledge and cue redundancy in deductive reasoning. *Memory & Cognition, 13*, 453-462.

Hoenkamp, E. (1989). 'Confirmation bias' in rule discovery and the principle of maximum entropy. In *Proceedings of the 11th Annual Conference of the Cognitive Science Society* (pp. 651-658). Hillsdale, NJ: Erlbaum.

Horwich, P. (1982). *Probability and evidence*. Cambridge, England: Cambridge University Press.

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

Jackson, S. L., & Griggs, R. A. (1990). The elusive pragmatic reasoning schemas effect. *Quarterly Journal of Experimental Psychology, 42A,* 353–373.

Jaynes, E. T. (1978). Where do we stand on maximum entropy? In R. D. Levine & M. Tribus (Eds.), *The maximum entropy formalism.* Cambridge, MA: MIT Press.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, England: Cambridge University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. N., & Byrne, R. M. J. (1992). Modal reasoning, models, and Manktelow and Over. *Cognition, 43,* 173–182.

Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, M. S. (1972). Reasoning and a sense of reality. *British Journal of Psychology, 63,* 395–400.

Johnson-Laird, P. N., & Wason, P. C. (1970a). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology, 1,* 134–148.

Johnson-Laird, P. N., & Wason, P. C. (1970b). Insight into a logical relation. *Quarterly Journal of Experimental Psychology, 22,* 49–61.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

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

Kirby, K. N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition, 51,* 1–28.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94,* 211–228.

Klayman, J., & Ha, Y. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 596–604.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge, England: Cambridge University Press.

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

Luttrell, S. P. (1985). The use of transformations in data sampling schemes for inverse problems. *Inverse Problems, 1,* 199–218.

MacFarland, D. (1977). Decision making in animals. *Nature, 269,* 15–21.

MacFarland, D., & Houston, A. (1981). *Quantitative ethology: The state-space approach.* London: Pitman Books.

MacKay, D. J. C. (1992). *Bayesian methods for adaptive models.* Unpublished doctoral dissertation, California Institute of Technology, Pasadena.

Manktelow, K. I., & Evans, J. St. B. T. (1979). Facilitation of reasoning by realism: Effect or non-effect. *British Journal of Psychology, 70,* 477–488.

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

Manktelow, K. I., & Over, D. E. (1990a). Deontic thought and the selection task. In K. J. Gilhooly, M. Keane, R. Logie, & G. Erdos (Eds.), *Lines of thought: Reflections on the psychology of thinking.* New York: Wiley.

Manktelow, K. I., & Over, D. E. (1990b). *Inference and understanding.* London: Routledge.

Manktelow, K. I., & Over, D. E. (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition, 39,* 85–105.

Manktelow, K. I., & Over, D. E. (1992). Utility and deontic reasoning: Some comments on Johnson-Laird and Byrne. *Cognition, 43,* 183–188.

Manktelow, K. I., & Over, D. E. (1993). *Rationality: Psychological and philosophical perspectives.* London: Routledge.

McArthur, L. Z. (1972). The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology, 22,* 171–193.

Moshman, D. (1978). Some comments on Bree and Coppens' "The difficulty of an implication task." *British Journal of Psychology, 69,* 371–372.

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

Oaksford, M. (1993). Mental models and the tractability of everyday reasoning. *Behavioral and Brain Sciences, 16,* 360–361.

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind and Language, 6,* 1–38.

Oaksford, M., & Chater, N. (1992). Bounded rationality in taking risks and drawing inferences. *Theory and 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., & Stenning, K. (1990). Connectionism, classical cognitive science and experimental psychology. *AI & Society, 4,* 73–90.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 835–854.

Ormerod, T. C., Manktelow, K. I., & Jones, G. V. (1993). Reasoning with three types of conditional: Biases and mental models. *Quarterly Journal of Experimental Psychology, 46A,* 653–677.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufman.

Platt, J. R. (1964). Strong inference. *Science, 146,* 347–353.

Pollard, P. (1985). Nonindependence of selections on the Wason selection task. *Bulletin of the Psychonomic Society, 23,* 317–320.

Popper, K. R. (1959). *The logic of scientific discovery.* London: Hutchinson.

Putnam, H. (1974). The "corroboration" of theories. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (Vol. 1, pp. 221–240). LaSalle, IL: Open Court.

Quine, W. V. O. (1953). Two dogmas of empiricism. In W. V. O. Quine (Ed.), *From a logical point of view* (pp. 20–46). Cambridge, MA: Harvard University Press.

Quine, W. V. O. (1959). *Methods of logic.* New York: Holt, Rinehart & Winston.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118,* 219–235.

Reich, S. S., & Ruth, P. (1982). Wason's selection task: Verification, falsification and matching. *British Journal of Psychology, 73,* 395–405.

Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review, 90,* 38–71.

Rips, L. J. (1990). Reasoning. *Annual Review of Psychology, 41,* 321–353.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension.* Hillsdale, NJ: Erlbaum.

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General, 110,* 101–120.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication.* Urbana: University of Illinois Press.

Siegel, S., & Castellan, N. J., Jr. (1988). *Non-parametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Simon, H. A. (1959). Theories of decision making in economics and behavioral science. *American Economic Review, 49*, 253–283.

Skilling, J. (Ed.). (1989). *Maximum entropy and Bayesian methods*. Norwell, MA: Kluwer Academic.

Smalley, N. S. (1974). Evaluating a rule against possible instances. *British Journal of Psychology, 65*, 293–304.

Staudenmayer, H., & Bourne, L. E. (1978). The nature of denied propositions in the conditional sentence reasoning task. In R. Revlin & R. E. Mayer (Eds.), *Human reasoning* (pp. 83–99). New York: Wiley.

Stich, S. (1985). Could man be an irrational animal? *Synthese, 64*, 115–135.

Stich, S. (1990). *The fragmentation of reason*. Cambridge, MA: MIT Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 423–469.

Valentine, E. R. (1985). The effect of instructions on performance in the Wason selection task. *Current Psychological Research and Reviews, 4*, 214–223.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*, 129–140.

Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmonsworth, Middlesex, England: Penguin.

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

Wason, P. C. (1969). Regression in reasoning. *British Journal of Psychology, 60*, 471–480.

Wason, P. C., & Golding, E. (1974). The language of inconsistency. *British Journal of Psychology, 65*, 537–546.

Wason, P. C., & Green, D. W. (1984). Reasoning and mental representation. *Quarterly Journal of Experimental Psychology, 36A*, 597–610.

Wason, P. C., & Johnson-Laird, P. N. (1970). A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology, 61*, 509–515.

Wason, P. C., & Johnson-Laird, P. N. (1972). *The psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology, 23*, 63–71.

Wiener, N. (1948). *Cybernetics*. New York: Wiley.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Newbury Park, CA: Sage.

Yachanin, S. A. (1986). Facilitation in Wason's selection task. *Current Psychological Research and Reviews, 5*, 20–29.