MARTIN JONES and ROBERT SUGDEN

POSITIVE CONFIRMATION BIAS IN THE ACQUISITION OF INFORMATION

ABSTRACT. An experiment is reported which tests for positive confirmation bias in a setting in which individuals choose what information to buy, prior to making a decision. The design – an adaptation of Wason's selection task – reveals the use that subjects make of information after buying it. Strong evidence of positive confirmation bias, in both information acquisition and information use, is found; and this bias is found to be robust to experience. It is suggested that the bias results from a pattern of reasoning which, although producing sub-optimal decisions, is internally coherent and which is self-reinforcing.

KEY WORDS: Positive confirmation bias, Selection task, Information acquisition

Traditionally, economics has assumed that economic agents are rational optimizers. This assumption has often been casually defended by means of the argument that, by repeated experience of market transactions, agents will learn to optimize. Recently, however, economists have begun to investigate and to theorize about the actual mechanisms by which individuals learn, and to ask whether these mechanisms induce learning trajectories which converge on optimizing behaviour (e.g. Roth and Erev, 1995; Börgers, 1996; Börgers and Sarin, 1997; Cubitt and Sugden, 1998). Our paper is a contribution to this larger enterprise.

Many psychologists have proposed that human reasoning is subject to *positive confirmation bias*. This is a tendency, when testing an existing belief, to search for evidence which could confirm that belief, rather than for evidence which could disconfirm it.[1] In general, both kinds of evidence are relevant for appraising the validity of a belief; there is bias if, relative to norms of valid reasoning, excessive effort is devoted to the search for confirming evidence. If positive confirmation bias is a fundamental property of the processes of inference and learning used by human beings, then we might expect it to impact on the decisions that economic agents make in relation to the acquisition of information. As a result, there might be systematic biases in economic learning; for example, an agent who

Theory and Decision **50:** 59–99, 2001.

repeatedly faces the same set of options might retain the false belief that a particular option was optimal, even after long exposure to evidence which, rationally interpreted, would indicate the contrary.

The primary objective of the research reported in this paper is to test for the existence of positive confirmation bias in a controlled experimental setting in which individuals choose what information to buy, prior to making a decision. The design of our experiment is modelled on Wason's (1968) *selection task*. This task, in various guises, is the paradigm most used by experimental psychologists when investigating positive confirmation. However, in the forms in which it has been used up to now, the selection task is not a decision-making problem in the economic sense.

Experimental subjects have simply been asked to say what information they would need to gather in order to be sure of the truth or falsity of a particular statement; they have not been asked to think about the costs of acquiring information or about the benefits of using it. This feature of the existing evidence may explain why economists have shown much less interest in positive confirmation bias than in many other experimentally-observed 'anomalies', which can be interpreted as violations of standard theories of decision-making. Our experiment reveals a pattern of information-gathering behaviour which contravenes the fundamental principles of Bayesian decision theory.

A related limitation of previous investigations of positive confirmation is that they do not reveal what use individuals make of information after they have gathered it. Existing evidence from selection tasks suggests that individuals seek certain kinds of information which, in the framework of a theory of rationality, is valueless. The implications of such behaviour for an economic theory of learning depend crucially on whether irrelevant information is simply ignored in subsequent decision-making or is treated as if it were relevant. The use to which irrelevant information is put also has implications for individuals' ability to learn by experience that such information is not worth collecting. Our experiment investigates the use made of information and the effect of experience on information-gathering and information-using behaviour. In the light of our findings, we shall suggest that positive confirmation bias may be robust to experience.

1. THE WASON SELECTION TASK: EXISTING THEORY AND EVIDENCE

The original selection task (Wason, 1968; Wason and Johnson-Laird, 1972) is deceptively simple in its design. A typical experiment uses a layout of four double-sided cards. Subjects are told that each card has a letter on one side and a number on the other, but they can see only the upper faces of the four cards. These show 'A', 'D', '4' and '7'. Each subject is asked to consider the following 'rule', as applied to the four cards: 'If a card has a vowel on one side, then it has an even number on the other side'. The instruction takes the form: 'Your task is to say which of the cards you need to turn over to find out whether the rule is true or false.' The two most common responses are the 'A' card alone, and the 'A' and '4' cards in combination. The correct answer to the question posed is, of course, the combination of 'A' and '7'. The frequently-chosen '4' card can provide no information which is relevant to the issue of whether the rule is true or false. Notice, however, that the 'A' and '4' cards are the ones that are capable of providing evidence which *confirms* the rule: by turning over either of these cards, the subject may find a card with a vowel on one side and an even number on the other. In contrast, the '7' card can only *disconfirm* the rule (i.e. by revealing a card which has a vowel on one side but not an even number on the other). In this sense, the evidence from the selection task can be interpreted as consistent with positive confirmation bias. This is Wason's own interpretation of his results.

From now on, we will use a more general notation to describe selection task experiments. We shall say that the cards have *labels* on each side. The *meanings* of these labels are the *propositions* p, q, ¬p and ¬q, where ¬ is the logical operator 'not'. We treat propositions (e.g. 'the card has a vowel on one side') as subject to the rules of logic. However, we treat the labels themselves (e.g. «A») as objects, such as strings of letters, which are distinct from their meanings. Such objects will be enclosed by guillemets (« ... »).

The subject is asked to test the truth or falsity of the *statement* «If [p], then [q]». Here the guillemets signify that the statement is to be understood as a string of words. The square brackets around p and q signify that in the statement actually presented to the subject, labels which mean p and q are used. The *meaning* of the statement

is the *proposition* p ⇒ q. The distinction between statements and propositions is important, because we need to be able to distinguish between the two statements ≪If [p], then [q]≫ and ≪If [¬q], then [¬p]≫, even though those statements have equivalent meanings.

The subject's response will be described by the set of cards which she opts to turn over. Usually, we shall identify the four cards simply by their upper faces. In this notation, the most common responses are {p} and {p, q} while the correct response is the set of *informative* cards {p, ¬q}. Occasionally, we shall use the notation <g, h> to denote a card whose upper face is g and whose lower face is h. (Here and throughout the paper, g, h∈S where S = {p, q, ¬p, ¬q}.) In this latter notation, a card whose upper face is g but whose lower face is unknown or unspecified will be denoted by <g, #>.

Since the publication of Wason's findings, there has been an explosive growth of literature on the subject. The replicability of Wason's original result is not in dispute, but how it should be interpreted remains a matter of debate. This debate has been informed by a large number of selection task experiments in which different modifications have been made to the original design; the theoretical problem has been to explain why some versions of the experiment induce the incorrect responses {p} and {p, q} while other versions facilitate the correct response {p, ¬q}. In the following subsections we outline some of the most important hypotheses that have been proposed to account for the evidence. A secondary objective of our experiment is to try to discriminate between these explanations.

1.1. Matching

Evans (1972) challenges the claim that the results of selection task experiments are evidence of positive confirmation bias. His theory proposes that there is a *matching bias*, such that subjects in the selection task experiment tend to choose whichever cards happen to be named in the statement to be tested, ignoring negations. Thus, faced with the statement ≪If a card has a *vowel* on one side, then it has an *even number* on the other side≫, subjects simply respond by choosing those cards which show vowels or even numbers. Matching is not interpreted as a form of reasoning, but as a mental processing fault resulting from an overload of the subject's cognitive and perceptive abilities. This hypothesis implies that subjects' responses

can be changed by rephrasing the statement, introducing negations but preserving its meaning. For example, if the statement is rephrased as ≪ If a card has a *vowel* on one side, then it does not have an *odd number* on the other ≫, subjects will tend to choose those cards which show vowels and odd numbers; this would be the correct response, but not chosen by virtue of its correctness. Subsequent research has confirmed that matching bias does occur (Evans and Lynch, 1973; Manktelow and Evans, 1979). We shall discriminate between positive confirmation and matching bias by looking for consistency between subjects' strategies for gathering information and their strategies for using it. Such consistency, if found, would suggest that subjects were engaging in reasoning rather than suffering from mental overload.

1.2. Realism

The original selection task was formulated in highly abstract terms. It has been suggested that the correct response might be facilitated by adding *thematic* content to the task, so that the statement is more readily intelligible to subjects. This can be done by making p and q refer to less abstract entities than letters and numbers, and by providing a *cover story* which accounts for the statement and gives some point to the selection task. The first experimentalists to investigate the effects of thematic content were Johnson-Laird, Legrenzi and Legrenzi (1972), who proposed that the crucial variable was the *realism* of the cover story, as viewed by the subject. Their findings supported the hypothesis that realistic cover stories facilitate the correct response. Our experiment tests the effects of using cover stories with different degrees of realism.

1.3. Pragmatic reasoning schemata

Following the work of Johnson-Laird et al., subsequent research has shown that while some cover stories tend to induce the correct response, others do not (e.g. Manktelow and Evans, 1979; Griggs and Cox, 1982; Reich and Ruth 1982). It now seems that realism and familiarity may not be the most important facilitating factors. A theory of *pragmatic reasoning schemata*, put forward by Cheng and Holyoak (1985), is increasingly gaining acceptance as an explanation of the effects of cover stories.

A pragmatic reasoning schema is a set of abstract, generalized, context-sensitive rules which can be applied to a particular class of problems. These rules are 'pragmatic' in the sense that they are non-logical; they are used in place of logic because of the latter's relative complexity. By setting a context for a selection task, thematic material can determine which schema (if any) the subject invokes. Cheng and Holyoak propose that there are *deontic* schemata of obligation and permission which are particularly effective in facilitating the correct response to the selection task. A proposition $p \Rightarrow q$ is an *obligation* if it takes the form 'If action p' is performed, then condition q' is obligatory' (e.g. 'If a person is drinking beer, then that person must be over 18').[2] Cheng and Holyoak propose that individuals use schemata which facilitate the correct response in the selection task *if the cover story is deontic*.

On this hypothesis, the realism and familiarity of the cover story are irrelevant; what matters is its deontic structure. Cover stories, however realistic and familiar, will not facilitate the correct response if the $p \Rightarrow q$ proposition is a *causal* relationship (e.g. 'If a person has drunk whisky, then that person has a high blood alcohol level') or a *neutral* (i.e. non-deontic, non-causal, non-tautological) material conditional (e.g. 'If a person in this room is male, then that person smokes'). There is now a considerable body of evidence that deontic cover stories facilitate the correct response, particularly if the subject is asked to adopt the viewpoint of the enforcer of a deontic rule (e.g. Cheng and Holyoak, 1985; Cosmides, 1989; Manktelow and Over, 1991). A possible explanation for this finding is that, to *enforce* a rule, one has to look for violations of it – that is, to look for disconfirmations. In contrast, when considering whether an «If [p], then [q]» statement *is true*, people may be predisposed to look for confirmations. Our experiment will test the differential effects of neutral, causal and deontic cover stories.

1.4. 'Bayesian' reasoning

The selection task, as usually formulated, is amenable to Bayesian analysis only in the most trivial sense. Whatever the subject's priors, he can be certain of the truth or falsity of the proposition $p \Rightarrow q$ by turning over the p and $\neg q$ cards. Provided that all relevant prior probabilities lie strictly between 0 and 1, he cannot have posterior

subjective certainty about the truth or falsity of the proposition *unless* he turns over both of these cards. Since he is asked to say which cards he *needs* to turn over to find out whether the statement is true or false, {p, ¬q} is the unambiguously correct response, irrespective of his (diffuse) priors.

However, some commentators have used Bayesian analysis to argue that behaviour in the selection task is guided by heuristics which, although inappropriate for that task, are well-adapted to many real-world problems. An interpretation of the selection task evidence along these lines is offered by Fischoff and Beyth-Marom (1983). A similar argument is presented by Klayman and Ha (1987). Klayman and Ha consider the heuristics that people might use to test hypotheses of the form 'p tends to be associated with q', which apply across a wide domain (for example: 'smoking is a cause of lung cancer'). Relevant information for assessing the truth of such a hypothesis can be found by sampling any of the cases p, q, ¬p, ¬q; but if the unconditional probabilities of p and q are low, samples of p (taking a sample of smokers and finding how many of them contract lung cancer) and q (taking a sample of people with lung cancer and finding how many of them are smokers) are more informative than samples of ¬p and ¬q. Thus, a heuristic which prompts people to test such hypotheses by sampling cases of p and q is well-adapted to many real problems.[3,4]

Notice that this argument does not imply that the choice of p and q *in Wason's selection task* is rational in the Bayesian sense. Nor does it imply that similar deviations from norms of rationality are absent in real-world decision-making. The argument offers an *explanation* of why positive confirmation bias occurs: the bias is a by-product of heuristics which, on the whole, work reasonably well. To this extent, the argument has little bearing on the design on experiments such as ours, which are designed to test whether positive confirmation bias exists.

However, these Bayesian lines of reasoning point to the potential significance of subjects' prior beliefs about probabilities. In particular, whether the frequently-chosen p card has more or less information content than the rarely-chosen ¬q card depends on subjects' priors. Our experiment is designed so as to achieve as much control as possible over subjects' priors.

2. EXPERIMENTAL DESIGN: PRINCIPLES

The principal objective of our experiment is to test for positive confirmation bias in a setting in which individuals make information-acquisition decisions which have real financial consequences for them. In this section, we explain the broad principles of the design; its practical implementation is described in Section 5.

The experiment uses a pack of double-sided cards; each card is labelled so that it has (strictly: has a label whose meaning is) p or ¬p on one side and q or ¬q on the other. The up/down orientation of each card is fixed during the experiment, and so there are effectively eight possible types of card <g, h>. For each subject and for each task, the total number of cards in the pack is the same; we denote this number m. The composition of the pack is determined by a random process, independently for each subject and for each task. This process uses a parameter $\alpha \in (0, 1)$, which is constant across subjects and tasks. The type of each of the m cards is determined independently; using s(<g, h>) to denote the probability that each card is of type <g, h>, the process is described by s(<p, q>) = s(<q, p>) = s(< ¬p, ¬q>) = s(< ¬q, ¬p>) = $\alpha/4$ and s(<p, ¬q>) = s(< ¬p, q>) = s(<q, ¬p>) = s(< ¬q, p>) = $(1 - \alpha)/4$.

It is useful to define a *contraposition function* f(.) such that f(p) = ¬q, f(q) = ¬p, f(¬p) = q, f(¬q) = p. We extend the domain of f(.) to the power set of S by defining f(A) = {f(g): g ∈ A} for all A⊆S. Notice that for all g, h: s(<g, h>) = s(<f[g], f[h]>). This property, which we shall call *contrapositive symmetry*, is important for our hypothesis tests. One of its implications is that, in a Bayesian analysis in which card frequencies are used as priors, there is no systematic difference between the information contents of the lower faces of the p and ¬q cards.

The subject inspects the pack before any cards are dealt from it. Four cards are then dealt at random, subject to the constraint that these cards are <p, #>, <q, #>, < ¬p, #> and < ¬q, #>. The subject is asked to consider the statement ≪Every card in the sample which is [p] is also [q]≫ or, for short, ≪Every [p] is [q]≫. She then chooses which if any of these cards to turn over; she has to pay a fixed *cost* per card turned over. After she has made this choice, the cards she has chosen are turned over (all together: it is not permitted to turn over one card and then, in the light of the information it

provides, to decide whether to turn over another). She then makes the *judgment* that the statement is 'true' or 'false'. Finally the remaining cards are turned over and she receives a fixed *reward* if and only if her judgment was in fact correct.

The cost and the reward are described to the subject in terms of 'points'. The subject starts the experiment (which may include more than one task) with an endowment of points sufficient to guarantee that she ends with a non-negative total. At the end of the experiment, she enters a lottery in which the prize is some fixed amount of money and in which the probability of winning that prize is proportional to the total number of points credited to her. Thus, if the subject is rational in the sense of expected utility theory (and given the trivial assumption that a money prize is preferred to nothing), she will seek to maximize the expected number of points scored in each task. More generally, the rule of maximizing expected points is implied by *any* theory of rational choice in which higher probabilities of preferred outcomes are preferred to lower probabilities of the same outcomes. Such preferences will be called *dominance-respecting*. As a normalization, we define the subject's *payoff* for a task as the score in points divided by the points value of the reward. We use c to denote the cost in points of turning over a card as a ratio of the points value of the reward. Thus, the subject loses c units of payoff for each card turned over and gains one unit of payoff if her judgment is correct.

This *binary lottery* incentive system has been widely used in experimental economics as a means of inducing risk-neutral preferences. In a recent paper, Selten, Sadrieh and Abbink (1999) have proposed the hypothesis that, in fact, experimental subjects are at least as risk-averse with respect to payoffs that are denominated in terms of lottery tickets as they are with respect to payoffs that are denominated in money; and they have presented supporting evidence. If this hypothesis is true, we cannot assume that subjects in our experimental design are risk-neutral with respect to points. But, as we explain in Section 4, our null hypotheses are independent of subjects' attitudes to risk. Thus, such a failure of risk neutrality would not confound our tests for positive confirmation bias.

The subject's task can be analysed according to Bayesian decision theory and, as a benchmark, this analysis is presented in Sec-

tion 3. However, it is not our objective to test the hypothesis that our subjects are rational Bayesians. Even for someone who understands the principles of decision theory, the expected payoffs of the various strategies can be worked out only by counting the different kinds of cards in the pack and then doing a certain amount of analysis and computation. It is safe to assume that most of our subjects would not have known how to work out an optimal strategy from the information at their disposal. The mechanism of dealing the cards from a pack which the subject has inspected is used to secure experimental control over subjects' information; we do not need to assume that subjects' subjective beliefs about the lower faces of the four cards correspond with the relative frequencies of the different types of cards in the pack.

Our objective is to test whether people's decision-making exhibits a particular systematic bias, namely positive confirmation bias. Accordingly, we need null hypotheses which permit the widest possible range of behaviour – rational or irrational – while excluding those kinds of behaviour that would result from positive confirmation bias. After the Bayesian analysis, we shall explain our null and alternative hypotheses.

3. BAYESIAN ANALYSIS

We define a *strategy* as the combination of a set of cards to be turned over and a rule which conditions the subject's judgment on the information she receives. We evaluate alternative strategies in terms of their expected payoffs. As explained in Section 2, it is an implication of Bayesian decision theory that a *rational* subject will maximize expected payoff in each task; since this is a theorem, not an empirical hypothesis, it remains true even if, in fact, the binary lottery system fails to induce risk neutrality.

First, we eliminate those strategies which can be shown not to maximize expected payoff for *any* composition of the pack and for *any* $c > 0$. In a Bayesian analysis, the probability that the statement is true is independent of the downward faces of the uninformative cards <q, #> and < ¬p, #>. If both the informative cards <p, #> and < ¬q, #> are turned over, the truth or falsity of the statement is known with certainty. If only one of the informative cards is turned

over and is found to disconfirm the statement (i.e. if the card is <p, ¬q> or < ¬q, p>), then the statement is certainly false. If only one informative card is turned over and it fails to disconfirm the statement, then the probability that the statement is true must be revised upwards (provided the prior probability was non-zero). From these propositions it is straightforward to deduce that if a strategy maximizes expected payoff, it must be one of the following five *qualitatively rational* strategies:

S_1: Turn over no cards; judge statement true.
S_2: Turn over no cards; judge statement false.
S_3: Choose {p}; judge statement true iff lower face of chosen card is q.
S_4: Choose {¬q}; judge statement true iff lower face of chosen card is ¬p.
S_5: Choose {p, ¬q}; judge statement true iff (lower face of p card is q and lower face of ¬q card is ¬p).

Which of these strategies is optimal depends on the composition of the pack and on the value of c. In our experiment, the parameter values are set at m = 100, α = 0.8, and c = 0.125. For these values, and for the *average* pack of cards in which the proportion of cards of each type <g, h> is equal to the prior probability s(<g, h>), it can be shown that the expected payoffs of strategies $S_1, \ldots, S_5$ are respectively 0.64, 0.36, 0.715, 0.715, and 0.75. In this sense, the *objectively optimal* strategy is S_5: since this requires {p, ¬q} to be chosen, our design can be thought of as a Bayesian analogue of the original selection task.

As these calculations show, the expected payoff from S_5 is only slightly greater than that of several other strategies. This is an unavoidable feature of our design. From their knowledge of the composition of the pack, subjects can assess the prior probability that the statement is true; so even if no cards are turned over, the judgment of a Bayesian subject must have at least a 0.5 probability of being correct, whatever the value of α. It is essential that subjects perceive there to be a significant cost to turning each card, since otherwise there would be no reason not to turn them all over. The presence of this cost reduces the expected payoff from turning over the two informative cards. In consequence, the expected payoff from turning over two cards cannot be very much greater than *both* the expected

payoff from turning over one card *and* that of turning over none. This property of expected payoffs would be problematic *if* the experiment were designed to test Bayesian rationality. But, to repeat: that is not our objective. Our objective to test for positive confirmation bias. Our null hypothesis is not Bayesian rationality; it is the absence of that bias.

4. THE PRINCIPAL HYPOTHESES TO BE TESTED IN THE EXPERIMENT

In this section we explain our approach to testing for positive confirmation bias, both in subjects' choices of cards and in their subsequent judgements about the truth or falsity of the statement.

4.1. Choice of cards

We work in a framework of *stochastic* choice, applied to a given subject, for given values of the parameters m, α and c, and for a given labelling of the cards. Interpreting p, q, ¬p, ¬q as the four cards, we define a function $\pi(.)$ from the power set of S to the interval [0, 1]. For each $A \subseteq S$ (i.e. for each set of cards which might be chosen), $\pi(A)$ is interpreted as a *decision probability*: it is the probability (assessed prior to the deal of the four cards) that the subject chooses A when testing the statement «Every [p] is [q]». Stochastic variation in choice is to be interpreted as resulting from errors or imprecision in individuals' preferences and beliefs (Loomes and Sugden, 1995) as well as from the effects of random variation in the composition of the pack of cards. By modelling choice as stochastic, we allow for the possibility that subjects' behaviour is partly random. Our tests look for patterns in that behaviour that cannot be explained by random variation.

Now consider the same subject, the same parameter values, and the same labelling of the cards, but a different statement. Specifically, the statement is «Every [p′] is [q′]», where p′ = f(p) and q′ = f(q). We use T′ to denote the task of testing this statement; the task of testing «Every [p] is [q]» is denoted by T. We define a function $\pi'(.)$ from the power set of S′ = {p′, q′, ¬p′, ¬q′} to [0, 1], specifying decision probabilities for task T′. We shall say that p is *isomorphic* with p′, q with q′ and so on. Similarly, two sets

$A \subseteq S$ and $A' \subseteq S'$ are isomorphic if A can be transformed into A' by substituting p' for p, q' for q, and so on. For example, suppose the cards are labelled «vowel» or «consonant» on one side and «even» or «odd» on the other, and that T is the task of testing «Every vowel card is an even-numbered card»; thus $\pi(\{p\})$ is the probability of choosing just the «vowel» card in T. Then T' is the task of testing «Every odd-numbered card is a consonant card» and $\pi'(\{p'\})$ is the probability of choosing just the «odd» card in T'.

Notice that the tasks T and T' are identical except for the wordings of the statements: the two statements have logically equivalent meanings. Thus, irrespective of the composition of the pack, and irrespective of the subject's attitude to risk, any theory which considers only the logical structure of decision problems and which ignores the way those problems are framed must imply that the decision probabilities associated with any given set of cards (defined in terms of their labels) is the same for both tasks. That is, for all $A \subseteq S$ and $A' \subseteq S'$ such that A and A' are isomorphic: $\pi(A) = \pi'(f[A'])$. We shall call this condition *cross-task contraposition neutrality*. If there is positive confirmation bias, however, cross-task contraposition neutrality will be violated. This is because the cards which can confirm the statements in the two tasks are different. (Using the example of the preceding paragraph: «vowel» and «even» can confirm the statement in T, while «odd» and «consonant» can confirm the statement in T').

Recall that the process which determines the composition of the pack satisfies contrapositive symmetry. This property implies that the isomorphism between T and T' extends to card frequencies. More precisely, for all g, h, g', h' such that g' is isomorphic with g and h' is isomorphic with h: $s(<g, h>) = s(<g', h'>)$. Thus, the only difference between A in task T and the isomorphic set A' in task T' is in respect of the labelling of the cards. Any theory which considers only the logical structure of decision problems and which ignores framing and labelling must imply that the decision probabilities associated with isomorphic sets are equal. That is, for all $A \subseteq S$ and $A' \subseteq S'$ such that A and A' are isomorphic: $\pi(A) = \pi'(A')$. Again, this result is independent of the subject's attitudes to risk. We shall call this condition *labelling neutrality*; violations of this condition are *labelling effects*.

One possible source of labelling effects is discussed by Oaksford and Chater (1994): subjects might use their background knowledge about the cover story, rather than their observations of the distribution of cards in the pack, to form priors. For example, take the case of the statement ≪Every person who is under 18 is drinking a soft drink≫. If the cover story is about a bar, the subject might believe that the majority of customers would be both over 18 and drinking alcohol, in which case the ≪under 18≫ card would be more informative than the ≪drinking alcohol≫ card.

If decision probabilities satisfy both cross-task contraposition neutrality and labelling neutrality, then for all $A \subseteq S$: $\pi(A) = \pi(f[A])$. We shall call this property *within-task contraposition neutrality*. Our principal test for bias looks at subjects' responses to a single task, and uses within-task contraposition neutrality as the null hypothesis.[5] The alternative hypothesis is that there are systematic divergences from this form of neutrality in the direction that would be consistent with positive confirmation bias. Roughly, our alternative hypothesis is that, other things being equal, the *potentially confirming* cards p and q are chosen more frequently than other cards, 'other things being equal' being interpreted in terms of within-task contraposition neutrality.

More precisely, for any given task, consider the following sets of cards: $A_1 = \{p\}$, $A_2 = \{q\}$, $A_3 = \{p, q\}$, $A_4 = \{p, q, \neg p\}$, $A_5 = \{p, q, \neg q\}$. Then $f(A_1) = \{\neg q\}$, $f(A_2) = \{\neg p\}$, $f(A_3) = \{\neg p, \neg q\}$, $f(A_4) = \{q, \neg p, \neg q\}$, and $f(A_5) = \{p, \neg p, \neg q\}$. The null hypothesis implies that $\pi(A_i) = \pi(f[A_i])$ for i = 1,. . . ,5, and hence that $\Sigma_i\ \pi(A_i) = \Sigma_i\ \pi(f[A_i])$. Notice that for each i, after eliminating cards which are common to both A_i and $f(A_i)$, A_i contains only potentially confirming cards while $f(A_i)$ contains only cards which are *not* potentially confirming. (A_1, ..., A_5 are the only sets of cards for which this is true.) Thus, our alternative hypothesis is that for each i, $\pi(A_i) > \pi(f[A_i])$, and hence that $\Sigma_i\ \pi(A_i) > \Sigma_i\ \pi(f[A_i])$.

This hypothesis is convenient because it applies to individual tasks, rather than requiring comparisons across tasks. However, a test of this hypothesis does not discriminate between positive confirmation bias and labelling effects. In order to discriminate between these two effects, we investigate some pairs of tasks which stand in the same relation to one another as do T and T′ in the preceding

theoretical discussion. Using A'_1, ..., A'_5 to denote sets of cards in T′ which are isomorphic with $A_1,\ldots,A_5$, it is easy to see that while *either* $\Sigma_i\ \pi(A_i) > \Sigma_i\ \pi(f[A_i])$ *or* $\Sigma_i\ \pi'(A_i) > \Sigma_i\ \pi'(f[A_i])$ could result from labelling effects, *both together* could not. Thus, to accept the alternative hypothesis in relation to both tasks is to conclude that there is a regularity in subjects' choices which (i) cannot be explained by any theory which considers only the logical content of statements and ignores the way they are framed, (ii) cannot be explained by labelling effects, but (iii) *can* be explained by positive confirmation bias.

4.2. True/false judgments

In Section 3 we described how a Bayesian agent would revise her beliefs about the truth or falsity of the statement in the light of information generated by turning over cards. We now ask in what respects a subject might be expected to deviate from those Bayesian judgments if she was influenced by positive confirmation bias.

We will say that a card <p, ¬q> or < ¬q, p>, if turned over, is a *disconfirmation*, while <p, q> and <q, p> are *confirmations*; <p, q> is an *informative* confirmation while <q, p> is an *uninformative* one. Recall that positive confirmation bias is an excessive tendency to seek confirmations when testing hypotheses; a person who is influenced by this bias is inclined to turn over the <p, #> and <q, #> cards because these might be the confirmations <p, q> and <q, p>. If confirmation-seeking is interpreted as part of a strategy for *testing* hypotheses, we should expect confirmations – even uninformative ones – to count in favour of a hypothesis. Thus, it is a natural extension of the theory of positive confirmation bias to propose that confirmations increase the subject's confidence in the truth of the statement.

This hypothesis implies systematic deviations from Bayesian rationality in response to the information revealed by turning over the <q, #> card (i.e. the card which has the potential to provide uninformative confirmations). Let $\tau(C)$ denote the probability that a given subject makes the judgement 'true', conditional on having turned over the set of cards C. (Here, a 'card' is defined by what is on each side, not just by its upper face.) Consider any two sets of cards C and D such that <q, p> ∈ C, <q, ¬p> ∈ D, and C\{<q,

p>} = D{<q, ¬p>}. Thus, the upper faces of the cards in C and D are identical, and both include the uninformative card <q, #>; the only difference between their lower faces is that C contains the confirmation <q, p> while D contains <q, ¬p>. We suggest that in such a case, a theory of positive confirmation should predict $\tau(C) > \tau(D)$. We shall test this hypothesis, using $\tau(C) = \tau(D)$ as our null.

5. EXPERIMENTAL DESIGN: DETAILS

The experiment was carried out at the University of East Anglia in Norwich. Subjects were recruited on the campus; most were students, coming from a wide range of course programmes. The 120 subjects took part in groups of up to twelve at a time. Each subject sat a screened computer workstation; there was no communication between subjects.

In the main part of the experiment, each subject faces a series of seven tasks, each of which has the general structure described in Section 2. Before starting these tasks, subjects are given full instructions about the nature of these tasks, about how points are scored, and about how points will be converted into chances of winning a prize. These instructions are given orally, in conjunction with a series of interactive instruction screens on each subject's workstation. This is followed by an example of the task; subjects work through this example with the help of further oral instructions. In composing the instructions, care was taken not to suggest that there was a right way to tackle any task, or to suggest that any particular strategy was to be preferred to any other.

Next, each subject answers three multiple-choice questions, designed to test her understanding of the task and of the scoring system. If a subject gives a wrong answer to any of these questions, further help is given. In the event, responses to these questions indicated a high level of understanding of the instructions.[6]

Each task is presented by means of a sequence of six screens. The first, *preliminary* screen presents the cover story and the statement. The cover story is in two parts. The *background* describes some set of 100 objects. Each object has two characteristics, each of which can take one of two values; these correspond with p, ¬p, q, and ¬q. A mechanism is described which explains how the characteristics

of each object have been written down on the two sides of a card, and how a sample of four of the 100 cards has been selected. For example, the background of the most abstract cover story is:

There is a collection of 100 objects each one of which is either 'Grue' or 'Bleen'. Also each object is either 'Smarge' or 'Lall'. Each object is described on a card with 'Grue' or 'Bleen' written on one side and 'Smarge' or 'Lall' written on the other. A sample of four cards is selected.

The *continuation* of the cover story introduces the statement in a way which characterizes it as neutral, deontic or causal. The statement itself is always expressed as a material conditional, irrespective of whether it has been explained in neutral, deontic or causal terms. For example, the neutral continuation of the 'Grue' cover story is:

Look at whichever cards you wish to test the statement:

Every object in the sample which is 'Grue' is also 'Smarge'.

The deontic continuation is:

There is a rule which requires that any object which is 'Grue' must also be 'Smarge'. To find out if the four objects in the sample are obeying the rule look at whichever cards you wish to test the statement:

Every object in the sample which is 'Grue' is also 'Smarge'.

All statements take the form ≪Every . . . in the sample which is . . . is also . . . ≫. We use this formulation, instead of the ≪If . . . , then . . . ≫ which has been used in many selection task experiments, for two reasons. First, the ≪If . . . , then . . . ≫ structure is ambiguous in ordinary language, and could be mistaken for the biconditional ≪If and only if . . . , then . . . ≫. Second, our formulation makes it explicit that the statement applies only to the sample of four cards, and not to any larger population (compare Section 1.4). No negations are used in any statement, so as to avoid any matching bias associated with negation (compare Section 1.1).

The second screen is the *browsing* screen. This screen represents a pack of 100 double-sided cards in random order, with the properties described in Section 2. The faces of these cards are labelled to correspond with the statement (e.g. ≪Grue≫, ≪Bleen≫, ≪Smarge≫ and ≪Lall≫). The pack is represented as being spread out on a table with the top card and the edges of the other cards showing. By using cursor keys, the subject can highlight the edge of any card; the front and back of that card are then displayed. After looking at as many

cards as she wishes, the subject moves on to the next screen. Thus, although the subject is not given any explicit information about the frequencies of different types of card in the pack, she has the opportunity to discover these frequencies by inspection if she wishes to do so.

This *shuffling* screen simply flashes the word 'shuffling' to signify that the program is re-randomizing the order of cards in the pack (while maintaining their up/down orientation). After a short interval, the *deal* screen is shown. This represents, in visual form, the random selection of <p, #>, <q, #>, < ¬p, #> and < ¬q, #> cards. Cards are seen to be 'dealt' from the top of the pack, with only their upper faces visible; the first card of each type to be reached is put into the standard four-card layout of the selection task. The four cards are placed side by side, working from left to right in the order in which they are dealt; this randomizes the relative positions of the p, q, ¬p and ¬q cards. As soon as the four cards have been dealt, the information on both sides of these cards is printed out. The subject does not see the printout until the end of the experiment, when she is given the opportunity to check that the undersides of the cards were determined before she made any decisions, and have not been changed since.

The fifth screen is the *choice* screen. The subject sees the upper faces of the four cards dealt out in the previous screen, and is asked to choose which cards to turn over. The statement to be tested is also displayed. At this stage, the subject can move freely between the choice screen and the preliminary screen, and so can refer to the cover story if she wishes. She must choose all the cards to be turned over before seeing any of their lower faces. The subject has been told that there is a cost of one point for every card turned over, and that the reward for a correct judgment about the truth or falsity of the statement is eight points. After she has made her choice, she moves on to the *question* screen. This screen displays both faces of the cards she has chosen to turn over, and the upper faces of the others. She is asked to judge whether the statement (which is displayed again) is true or false.

In performing the first six tasks, the subject sees only the five screens we have described so far. After making her judgment for the sixth task, the subject moves on to a series of six *answer* screens,

one for each of the previous tasks. The answer screen for a task displays both sides of all four cards, along with the statement and the subject's true/false judgment. The text on the screen tells the subject whether her judgment about the truth or falsity of the statement was right or wrong and if it was wrong, the reason why it was wrong (in the form 'The card with ... on the front had ... on the back').

Notice that this feature of the design implies that the subject has no feedback on the correctness of her judgments until she has completed the six tasks. By not providing such feedback, we ensure that a subject's responses to later tasks (in the set of six) are not influenced by her *success* in earlier tasks. However, some degree of cross-task learning is clearly possible, even in the absence of feedback on the correctness of judgments.[7] This is an unavoidable cost of a design in which each subject performs several tasks. Such learning, if it occurs, reduces the discriminatory power of our cross-task tests. But because the order of tasks is randomized (see Section 6), learning cannot induce systematic confounding effects.

The seventh task is a repeat of whichever task the subject faced first. In this task, the subject works through all six screens in sequence. By using this repeated task, our design allows us to investigate whether a subject's behaviour is influenced by feedback about the outcomes of her previous choices and judgments.

The subject begins the experiment with an endowment of 32 points. Thus, the total score for the experiment must lie in the range from 4 to 88; the objectively optimal strategy S_5 guarantees a score of 74. At the end of the experiment, each subject draws a lottery ticket from a set of tickets numbered from 1 to 100; if the ticket number drawn is less than or equal to the number of points scored, the subject wins £10.

6. COVER STORIES AND RANDOMIZATIONS

All the tasks in the experiment have the same logical structure, with the same parameter values m = 100, α = 0.8 and c = 0.125. They differ only in respect of the labelling of the cards, the cover stories, and the wording of the statement.

The cover stories are presented in full in the Appendix. There are six basic cover stories, which we shall identify by the objects

Table 1. Tasks used in the experiment

Cover story	p	q	Schema invoked	Form of statement	Realism of story
1: Relatives	London	Los Angeles	neutral	variable	realistic
2: Drinkers	alcohol	over 18	deontic	variable	realistic
3: Rivers	acidic	sick	causal	variable	realistic
4: Objects	grue	smarge	variable	standard	abstract
5: Diners	haddock	gin	variable	standard	fanciful
6: Meters	outside	red	variable	standard	fanciful

which the cards represent. Three of these – Relatives, Drinkers and Rivers – are *realistic* in the sense that they refer to concrete, readily understandable although necessarily stylized relationships. The Relatives story is neutral, the Drinkers story is deontic, and the Rivers story is causal. Each of these stories is used in two different but isomorphic treatments: the statement in the *contraposed* treatment is the contrapositive of that in the *standard* treatment. This feature of the design allows us to control for labelling effects.

The backgrounds of the other three cover stories – Objects, Diners and Meters – are written so that each story can be continued in different ways. Each of these stories has three alternative continuations – neutral, deontic and causal – which are used in different treatments. However, for a given cover story, the statement is the same in all three treatments. The object of this part of the design is to allow a controlled test of Cheng and Holyoak's (1985) hypothesis about pragmatic reasoning schemata, in the context of our design (see Section 1.3). According to that hypothesis, correct reasoning in selection tasks is facilitated if the cover story has a deontic structure, irrespective of the realism of that story. We deliberately did not try to make these three cover stories realistic: the Objects story is *abstract* (it uses predicates which have no meaning in the English language) while the other two are *fanciful*. Cheng and Holyoak (1989) use similarly fanciful cover stories when testing their hypothesis.

The tasks used in the experiment are summarized in Table 1. In the 'schema invoked' column, 'variable' signifies that neutral,

deontic and causal versions of the cover story are used in different treatments. In the 'form of statement' column, 'variable' signifies that standard and contraposed treatments are used in different treatments. (For these tasks, the entries under 'p' and 'q' in the Table refer to the standard treatment.) Each subject faces six different tasks, presented in random order, followed by a repeat of the first task.

Leaving aside the seventh task, each subject faces one task involving each of the six cover stories. For the *variable statement* tasks (cover stories 1 to 3), each subject either faces all three tasks in the standard treatment or faces them all in the contraposed treatment. For the *variable schema* tasks (cover stories 4 to 6), each subject faces one task in the neutral treatment, one in the deontic treatment, and one in the causal treatment. Within these constraints, subjects are assigned randomly to treatments.

Each instance in which a subject faces a task will be called a *case*. The experiment generates data for 840 cases. These cases can be partitioned into seven sets of 120 cases each, such that each set i = 1,. . . ,7 contains those cases in which a subject faces the ith task in her series of seven tasks. Notice that our randomization procedures ensure that each of these sets contains a random sample of the entire set of tasks and treatments. Thus, we can measure the effect of increases in experience by comparing aggregated responses for different values of i; we shall call i the *level of experience*. Notice also that, for every given cover story and treatment, subjects facing this task for the first time are randomly distributed over experience levels 1,. . . ,6. Thus, when making comparisons across tasks it is legitimate to aggregate across experience levels.

7. RESULTS: POSITIVE CONFIRMATION BIAS IN THE CHOICE OF CARDS

Table 2 presents some summary statistics about subjects' choices of cards. Each row of data in this table represents an aggregation across all 120 subjects. Each of the first six rows aggregates across the different treatments for a given cover story, while each of the next three rows aggregates across the three variable-schema cover stories for a given schema. To avoid double-counting, the data for the

Table 2. Cards turned over: summary stasistics

Task	n	Frequency of choices: sets of cards		Frequency of choices: individual cards				Test of differences between rows
		objectively optimal	qualitatively rational	p	q	¬p	¬q	χ^2
Relatives	120	26	71	94	43	7	41	2.90
Drinkers	120	67	101	92	18	2	77	46.75**
Rivers	120	29	69	79	46	14	51	7.33
Objects	120	19	68	86	46	13	39	n/a
Diners	120	34	72	91	45	7	52	9.89
Meters	120	35	75	90	43	7	47	8.05
neutral	120	24	66	86	50	12	38	n/a
deontic	120	33	70	91	43	8	49	4.04
causal	120	31	75	90	41	7	51	3.67
1st	120	27	78	74	37	9	39	n/a
2nd	120	28	80	88	36	6	40	5.56
3rd	120	31	73	89	43	8	50	5.99
4th	120	40	76	90	40	12	58	6.29
5th	120	42	71	97	45	7	60	12.79*
6th	120	42	78	94	40	8	60	15.73**
7th	120	41	67	104	49	10	64	17.88**
Total, excluding 7th	720	210	456	532	241	50	307	
(percent)		(29.2)	(63.3)	(73.9)	(33.5)	(6.9)	(42.6)	

* denotes significance at 95% confidence level.
** denotes significance at 99% confidence level.

different types of task refer only to the first six tasks faced by each subject. (Throughout our analysis of the results, we shall ignore all data relating to the seventh task, except when explicitly considering the effects of experience.) Each of the following seven rows of the table aggregates across all tasks for a given level of experience.

The first column of data shows the number of subjects who made the objectively optimal choice {p, ¬q}.[8] The second column shows the number who made qualitatively rational choices, i.e. who chose neither of the uninformative cards. The third column shows the number who chose the p card (either alone or in conjunction with other cards); the next three columns show corresponding numbers for the other cards. The final column shows a test statistic which will be explained later.

Table 3 shows the complete distribution of responses, broken down by type of task and by level of experience. Thus, the second column of data shows the frequency with which ∅ was chosen, the third shows the frequency with which {p} was chosen, the seventh shows the frequency with which {p, q} was chosen, and so on. The final two columns of Table 3 show the aggregations that are relevant for our tests of positive confirmation bias. The penultimate column shows the number of subjects who chose one of the sets $A_1, \ldots, A_5$, as defined in Section 4.1 (i.e. the sum of the columns marked by *). The final column shows the number who chose one of the isomorphic sets $f(A_1), \ldots, f(A_5)$ (i.e. the sum of the columns marked by #).

Looking at the column totals in Tables 2 and 3, it can be seen that the most frequent response is {p, ¬q}, i.e. the set of informative cards; this accounts for 29 percent of all responses. Considering individual cards, the cards most frequently chosen are p (74% of responses) and ¬q (43% of responses). Thus, *modal* responses are rational in the Bayesian sense. However, the overall distribution of responses shows a clear pattern that is consistent with positive confirmation. After {p, ¬q}, the most frequent responses are ∅ (18%), {p, q} (18%), {p} (14%), and {p, q, ¬q} (9%). Of these, {p, q} and {p} are the classic positive-confirmation responses, while {p, q, ¬q} seems to represent a mix of Bayesian rationality and positive confirmation. The uninformative but potentially confirming q card is chosen in 33 percent of cases.

Recall that the null hypothesis of within-task contraposition neutrality implies that the expected frequency of A_i responses is equal to that of $f(A_i)$ responses, while positive confirmation bias implies that former are more frequent than the latter. A glance at Table 3 is enough to show that this null hypothesis is decisively rejected. For

Table 3. Cards turned over

Task	n	Frequency of choice of set of cards																	
			*	*	#	#	*					#	*	*	#	#			
		no cards	p	q	¬p	¬q	p q	p ¬p	p ¬q	q ¬p	q ¬q	¬p ¬q	p q ¬p	p q ¬q	¬p ¬q	q ¬p ¬q	all cards	A_i sets (see text)	$f(A_i)$ sets (see text)
Relatives:																			
standard	63	5	15	0	0	1	14	3	13	1	0	1	0	10	0	0	0	39	2
contraposed	57	15	9	3	0	0	13	1	13	0	0	0	0	2	1	0	0	27	1
Drinkers:																			
standard	63	7	5	2	0	4	4	1	34	1	0	0	0	5	0	0	0	16	4
contraposed	57	13	5	1	0	0	4	0	33	0	0	0	0	1	0	0	0	11	0
Rivers:																			
standard	63	6	11	2	0	2	8	0	18	2	1	3	1	8	0	0	1	30	5
contraposed	57	17	4	2	2	0	11	0	11	2	2	0	1	3	0	0	2	21	2
Objects:																			
neutral	41	8	6	0	0	0	14	0	4	1	0	1	1	3	1	0	2	24	2
deontic	32	9	6	0	1	1	5	0	6	0	0	1	1	2	0	0	0	14	3
causal	47	7	11	0	1	1	10	0	9	0	1	1	0	4	0	1	1	25	4
Diners:																			
neutral	38	4	7	2	1	0	10	0	8	2	0	1	0	3	0	0	0	22	2
deontic	42	10	4	0	0	0	6	0	14	0	0	0	2	6	0	0	0	18	0
causal	40	8	5	0	0	0	6	1	12	0	1	0	0	7	0	0	0	18	0

Table 3 (continued)

Task	n	Frequency of choice of set of cards																	
		no cards	* p	* q	# ¬p	# ¬q	* p q	p ¬p	p ¬q	q ¬p	q ¬q	# ¬p ¬q	* p q ¬p	* p q ¬q	* p ¬p ¬q	# q ¬p ¬q	all cards	A_i sets (see text)	f(A_i) sets (see text)
Meters:																			
neutral	41	12	5	1	0	0	6	0	12	1	0	0	1	3	0	0	0	16	0
deontic	46	4	7	2	0	0	11	1	13	1	0	0	1	6	0	0	0	27	0
causal	33	8	4	0	0	0	7	0	10	1	0	0	0	2	1	0	0	13	1
1st	120	33	15	4	2	3	22	2	27	3	0	1	0	7	0	0	1	48	6
2nd	120	29	22	0	1	1	25	1	28	0	0	1	2	9	1	0	0	58	4
3rd	120	18	20	3	0	4	26	1	31	2	1	2	0	9	1	1	1	58	8
4th	120	22	14	4	0	0	17	1	40	2	0	2	2	11	1	0	4	48	3
5th	120	14	14	2	1	1	25	1	42	0	3	2	3	12	0	0	0	56	4
6th	120	17	19	2	1	0	14	1	42	5	1	0	1	17	0	0	0	53	1
7th	120	10	16	1	0	0	25	1	41	2	3	0	1	14	3	0	3	57	3
Total, excluding 7th	720	133	104	15	5	9	129	7	210	12	5	8	8	65	3	1	6	321	26

Note: A_i sets are marked by *; f(A_i) sets are marked by #.

every type of task and for every level of experience, the combined frequency of A_i responses is greater than that of $f(A_i)$ responses. In every case this difference is significant at the 99% confidence level (for a one-tailed test based on the binomial distribution). Notice that in the variable statement tasks, the standard and contraposed treatments produce the same massive asymmetry between A_i and $f(A_i)$ responses. Thus, the asymmetries in our data cannot be explained by labelling effects. There is, then, overwhelming evidence that subjects' information-gathering decisions are systematically biased in favour of information which is potentially confirming.

Despite the strength of this evidence of bias, behaviour seems to have been closer to Bayesian rationality in our experiment than in many selection task experiments. A meta-analysis of 34 non-thematic selection task experiments has shown that the individual cards p, q, ¬p, ¬q were chosen with frequencies 0.89, 0.62, 0.16 and 0.25 (Oaksford and Chater, 1994, p. 613), as compared with 0.74, 0.34, 0.07 and 0.43 in our experiment. We are reluctant to read too much into such comparisons, which are not subject to experimental control. But it may be worth saying that our design has various properties which could facilitate Bayesian rationality. One is the use of thematic material: it is not uncommon for {p, ¬q} to be the modal response to selection tasks when such material is used (e.g. Johnson-Laird et al, 1972; Cheng and Holyoak, 1985). Another is the presence of financial incentives, which might induce subjects to take greater care over their responses. A third is the framing of the task in terms of a cost *per card turned over*. This might prompt subjects to think about each card separately, and in so doing, to recognize which cards do and do not have information content. Finally, our design requires subjects to *use* the information that they collect by turning over cards. By thinking about how different kinds of information should be used, subjects might come to see which cards are capable of providing useful information.

8. RESULTS: COMPARISONS ACROSS CONDITIONS AND EXPERIENCE LEVELS

We have outlined various theories which offer reasons to expect differences in the extent of positive confirmation bias across tasks

and treatments. We test for such differences by comparing distributions of responses between types of task. For these tests, we classify responses into six categories – the five most common responses, i.e. {p, ¬q}, ∅, {p, q}, {p}, {p, q, ¬q}, and 'other'. Taking the most abstract cover story, Objects, as our datum, we test whether the distribution of responses for each other cover story is significantly different from that datum. Similarly, aggregating across the three variable-schema tasks, we take the neutral treatment as our datum and test whether the distributions of responses for deontic and causal treatments are significantly different from that. We also carry out similar tests for the effects of experience; here, the first task faced is the datum. Each test uses data from all 120 subjects. The final column of Table 2 reports the relevant χ^2 statistics (5 degrees of freedom; critical value for 95% confidence is 11.1).

We find no evidence to support Cheng and Holyoak's hypothesis that responses are different depending on whether a given statement is interpreted as a neutral, deontic or causal relationship. Nor do we find any general tendency for responses to be different depending on whether the cover story is abstract, fanciful or realistic. It is possible that the absence of significant cross-task differences is the product of cross-task learning (see Section 5). Nevertheless, our results show that responses *can* be affected by the cover story: compared with any of the other stories, Drinkers generates a quite different pattern of responses. (For the comparison between Drinkers and Objects, $\chi^2 = 46.75$ – an overwhelmingly significant difference.) On every criterion – the frequency of {p, ¬q} responses, the frequency of qualitatively rational responses, the frequency with which each individual card is chosen – responses to the Drinkers story are closer to Bayesian rationality and show less positive confirmation bias. Other researchers have used cover stories about under-age drinking in thematic selection tasks, and they too have found that this story facilitates the {p, ¬q} response (e.g. Griggs and Cox, 1982; Cheng and Holyoak, 1985).

Why this particular cover story has such a strong effect is not clear. It is reasonable to suppose that, for the young adults who make up the bulk of our subject pool, rules about under-age drinking are both familiar and salient. But if familiarity were crucial, one would expect our other realistic cover stories to have had at least

Table 4. Consistency between first and seventh task

	Response to first task						Total
Response to seventh task	Ø	{p}	{p, q}	{p,¬q}	{p,q,¬q}	other	
Ø	4	5	9	7	2	6	33
{p}	1	4	2	6	1	1	15
{p, q}	2	3	10	3	4	0	22
{p, ¬q}	0	1	0	21	3	2	27
{p, q, ¬q}	0	0	0	3	4	0	7
other	3	3	4	1	0	5	16
Total	10	16	25	41	14	14	120

some facilitating effect relative to Objects. It may be significant that Drinkers is a deontic cover story, but the results from our variable-schema tasks suggest that a deontic cover story is not a sufficient condition for facilitation. Since it would be futile to try to generalize from a sample of one, all we can safely say on the basis of our evidence is that some cover stories do facilitate the objectively optimal response, and that Drinkers is one such story.

The chi-squared tests for differences in responses across experience levels within the experiment show that some process of learning is going on: as the experience level increases, the distribution of responses becomes more and more different from the level 1 distribution. Surprisingly, however, there is no discontinuity between the sixth and seventh task, corresponding with subjects' receipt of feedback about the outcomes of the first six tasks. (For the comparison between the distributions of responses to the sixth and seventh tasks, $\chi^2 = 5.84$; the critical value for 95% confidence is 11.1).

Although it is quite clear that subjects are learning *something*, what they are learning is not so obvious. The frequency of the objectively optimal {p, ¬q} response increases with experience, but there is no decline in the frequency of positive-confirmation responses, as measured by choice of the q card or by A_i card combinations. The main decline is in the frequency of Ø responses.

Table 4 provides additional data about the effect of experience, cross-classifying the 120 subjects by their card choices in the first and seventh tasks. (Recall that for each subject, the seventh task is a repeat of the first.) Since the numbers in many of the cells of this table are quite small, we must be cautious is drawing strong conclusions from it; but some features of these data are worth noting.

We shall say that a particular response is *stable* to the extent that subjects who make that response in the first task make the same response in the seventh. A high degree of stability suggests that the response in question tends to generate reinforcement, that is, feedback which supports the subject's belief in the appropriateness of the response. The most stable response is the objectively optimal {p, ¬q}, which is repeated in 21 out of 27 cases.[9] Notice, however, that {p, q} and {p, q, ¬q} also have relatively high stability; of the 29 subjects who chose one or other of these responses in the first task, 18 still chose one or the other in the seventh task, while only 6 switched to the objectively optimal reponse. The implication seems to be that subjects who initially turn over the q card do not easily learn not to do so.

In contrast, Ø has low stability. Of the 33 subjects who chose Ø in the first task, only 4 made the same choice in the seventh, while 17 switched to responses which involved turning over at least one uninformative card. This suggests that, even though Ø is a qualitatively rational response, many of the subjects who chose it were not Bayesian reasoners. We shall defer further discussion of the effects of experience until we have looked at how subjects use information.

9. RESULTS: POSITIVE CONFIRMATION BIAS IN TRUE/FALSE JUDGEMENTS

Table 5 cross-classifies the 720 non-repeat cases by the information possessed by the subject after turning over any cards and by the subject's true/false judgment. The rows represent a breakdown of cases in terms of the information content of the informative cards <p, #> and < ¬q, #>. The columns present a breakdown in terms of the uninformative cards turned over, and the true/false judgment. The first pair of columns refer to cases in which neither of the uninformative cards was turned over. The second pair of columns refer to cases in

Table 5. True/false judgments

Informative cards turned over	Uninformative cards turned over, and judgment made							
	neither <q, #> nor <¬p, #>		<q,p>		<q,¬p>		<¬p, #> but not <q, #>	
	true	false	true	false	true	false	true	false
1. none	31	102	16	5	2	4	4	1
2. <p, q> only	65	16	85	6	13	7	6	1
3. <¬q, ¬p> only	3	4	2	1	0	0	7	0
4. <p, ¬q> only	0	23	1	20	0	5	0	0
5. <¬q, p> only	0	2	0	2	1	0	0	1
6. <p, q> and <¬q, p>	1	38	1	4	0	8	0	1
7. <p, ¬q> and <¬q, p>	0	11	0	3	0	0	0	1
8. <p,¬q> and <¬q,¬p>	1	22	0	8	0	2	0	1
9. <p, q> and <¬q,¬p>	134	3	36	0	6	3	0	0
Total	235	221	141	49	22	29	17	6
(percent true)	(51.5)		(74.2)		(43.1)		(73.9)	

which <q, #> was turned over and found to be the uninformative confirmation <q, p>. The third pair of columns refer to cases in which <q, #> was turned over and found to be <q, ¬p>. The final pair of columns contain the few residual cases, i.e. those in which < ¬p, #> was turned over but <q, #> was not.

One striking feature of the data is evident from rows 4 to 8: subjects almost always recognized the significance of a disconfirmation if they found it (the judgment 'false' was made in 152 of the 157 cases in which a disconfirmation was found). The data in row 9 are equally striking: when subjects had sufficient evidence to deduce that the statement was true, they again almost always made the correct judgment (176 cases out of 182). We shall say that a judgment is *deductively correct* if its truth can be established by logical deduction from the information available, and *deductively incorrect* if its falsity can be so established. Our data show that subjects – even those who turned over uninformative cards – rarely made deductively incorrect judgments.

Table 6. Effects of information on judgment

Information	Number of subjects judging:		χ^2
	true	false	
<p, q>	347	87	
<p, ¬q>	2	96	215.1**
<q, p>	141	49	
<q, ¬p>	22	29	17.7**
<¬p, q>	7	7	
<¬p, ¬q>	27	9	2.9
<¬q, p>	3	71	
<¬q, ¬p>	189	44	142.4**

*denotes significance at 95% confidence level.
**denotes significance at 99% confidence level.

At the least, these data provide strong evidence that subjects understood the tasks they faced and applied some form of reasoning to them. This suggests that the tendency to choose potentially confirming cards was not due to matching bias (see Section 1.1). We should not immediately infer that subjects actually went through the mental processes of deducing the truth or falsity of the statement from the available information. But clearly, whatever subjects' mental processes were, they were highly effective in those cases in which the statement's truth or falsity was in fact deducible.

In contrast, some subjects seem to have been rather unsuccessful in assessing the prior probability that the statement was true. Of those subjects whose response was Ø, 77 percent made the judgment 'false', while objectively (for the average pack) the prior probability that the statement is true is 0.64. We speculate that these subjects did not make systematic use of the opportunity to inspect the pack and so did not realize that cards with p on one side were very likely to have q on the other. The marked decline in the frequency of Ø responses over the course of the experiment suggests that these subjects gradually learned this property of the pack.

We can test for positive confirmation bias in subjects' judgments by looking at those subjects who turned over the q card, and at how their true/false judgments were affected by what they found. Table 5 shows that the statement was judged 'true' by 74 percent of the 190 subjects who found the confirmation <q, p>, but by only 43 percent of the 51 subjects who found <q, ¬p>. This difference is significant at the 99% confidence level (χ^2 = 17.7, one degree of freedom).[10] The implication is that, for subjects who turn over the q card, confidence in the truth of the statement is increased if the uninformative confirmation <q, p> is found. This is exactly what a theory of positive confirmation bias would predict (see Section 4.2). Table 6 reports this test, and the corresponding tests for each of the other three cards. As one would expect, subjects' judgments are strongly affected by what they find if and when they turn over the p and ¬q cards. No significant effect on judgments can be found for the ¬p card.

It seems, then, that many subjects are using reasoning processes which generate deductively correct judgments with a high degree of reliability, but which nevertheless are affected by positive confirmation bias at both the information-acquisition and the information-using stages. We now offer some tentative suggestions as to what those processes might be.

10. AN INTERPRETATION OF THE EVIDENCE

Consider the following class of *confirmation strategies*. Each such strategy can be described by a pair (A, v) where A⊆S, N(A) is the number of elements in A, and v ∈ {0, 1, . . . , N(*A*) + 1}. A person who follows (A, v) turns over the set of cards A and then makes the judgment 'true' if she finds *both* no disconfirmation *and* at least v confirmations. Otherwise, she makes the judgement 'false'. The five qualitatively rational strategies can be represented as confirmation strategies: S_1 as (∅, 0), S_2 as (∅, 1), S_3 as ({p}, 0) or ({p}, 1), S_4 as ({¬q}, 0), and S_5 as ({p, ¬q}, 0) or ({p, ¬q}, 1).

Now consider four confirmation strategies which are *not* qualitatively rational: S_6 = ({p, q}, 1), S_7 = ({p, q}, 2), S_8 = ({p, q, ¬q}, 1) and S_9 = ({p, q, ¬q}, 2). Notice that each of these strategies turns over a set of cards which was frequently chosen by our subjects,

and which includes the uninformative card q; each strategy also has positive confirmation bias at the judgment stage in that it fails to discriminate between informative and uninformative confirmations; but each makes the judgment 'false' whenever a disconfirmation is found. All these strategies are sub-optimal by virtue of incurring costs to acquire objectively valueless information, but their judgments are usually correct. (For the average pack of cards, the probability of being correct is 0.84 for S_6, 0.744 for S_7, 1.0 for S_8, and 0.872 for S_9; for comparison, the corresponding probabilities are 0.64 for S_1, 0.36 for S_2, 0.84 for S_3 and S_4, and 1.0 for S_5.) As a way of organizing our data, we offer the hypothesis that most subjects followed one or other of the strategies $S_1, \ldots, S_9$. We invite the reader to verify that the data in Tables 3 and 5 are consistent with this hypothesis.

A subject who uses one of the q-choosing strategies S_6, ..., S_9 does not make any distinction between <p, q> and <q, p> at the judgment-making stage. We speculate that such subjects are unaware of any difference between these two items of information: each is mentally recorded simply as 'a confirmation'. Recall that, until after the sixth task, subjects receive no feedback on the correctness of their judgments; but in performing each task, they discover what is on the lower faces of the cards they turn over. We suggest that turning over any particular card is psychologically reinforced to the extent that this action generates information which serves as an input to the subject's judgment-making process.

For a subject who recognizes the meaning of disconfirmations when she sees them and who treats confirmations as evidence that the statement is true, the p card always gives reinforcement and the q and ¬q cards sometimes do. This pattern of reinforcement may explain why the responses {p, q} and {p, q, ¬q} had relatively high stability, despite their involving turning over an uninformative card. Similarly, it may explain why the frequency with which each of the cards p, ¬q and (to a lesser extent) q was chosen increased with subjects' experience.

As we have said, we were surprised that the feedback provided between the sixth and seventh task had no significant effect on behaviour. In retrospect, however, this observation is perhaps explicable. Feedback about the correctness of judgments may have little impact

on subjects who are already using strategies which generate correct judgments with high probability. By the sixth task, most subjects (95 out of 120) were turning over at least one informative card. Confirmation strategies which have this property *do* usually produce correct judgments. Thus, subjects do not easily learn that the q card has no information value. In effect, that card has a parasitic relationship with the genuinely informative p card. Provided that the p card is turned over, confirmations *in general* are informative; for a subject who turns over both p and q and who does not distinguish between <p, q> and <q, p>, turning over q *appears* to generate useful information for a judgment-making process which is generally successful.

11. CONCLUSIONS

We draw three main conclusions from our research. First, we have found strong evidence of positive confirmation bias in the acquisition of information. This bias has been found in many previous psychological experiments, but as far as we know, our experiment is the first to use an incentive-compatible design in which subjects have to pay to acquire information. The bias that we have found is incompatible with all recognized versions of Bayesian decision theory.

Second, we have found a new form of positive confirmation bias in the use that is made of information: information which is interpreted as confirming a hypothesis increases subjects' confidence in the truth of the hypothesis, even if, viewed in a Bayesian perspective, that information has no value. This new finding throws valuable light on positive confirmation. It seems that positive confirmation bias is not a simple error; rather, it is a manifestation of a pattern of reasoning which, although producing sub-optimal decisions, is internally coherent.

Third, our findings suggest that positive confirmation bias may have a considerable degree of robustness to experience. More precisely, it seems that individuals can learn the value of looking for potentially disconfirming evidence, but that they persist in seeking confirmations which have no information value. We have speculated that this persistence of apparent irrationality might be explained

by a reinforcement theory of learning: given the conceptual framework within which individuals are working, confirmation-seeking strategies are reinforced.

We began this paper by locating it as a contribution to a larger enterprise – that of studying the mechanisms by which individuals learn from their experience, and of investigating whether these mechanisms tend to induce the optimizing behaviour that economics has traditionally assumed. Recognizing that what we are about to say may itself be evidence of the bias we have been studying, we believe that the results of our work provide further confirmation of the value of this enterprise.

ACKNOWLEDGMENTS

Robin Cubitt played an important part in the design of the experiment reported in this paper. We also thank Paul Anand, Alistair Munro, Chris Starmer and two anonymous referees for advice. Martin Jones's work was supported by a studentship from the Economic and Social Research Council (ESRC). Robert Sugden's work was supported by the ESRC research programmes on Economic Beliefs and Behaviour (award L 122 251 024) and Risk and Human Behaviour (award L 211 252 053) and by the Leverhulme Trust.

APPENDIX: THE COVER STORIES

1. Relatives

A survey is taken of 100 people in Los Angeles, Seattle, London and Norwich who have relatives living in other cities. Each person in the survey living in Britain has relatives in Los Angeles or Seattle and each person living in America has relatives in Norwich or London. No one has relatives in more than one city. The details of the survey are written down on report cards by putting the city each person lives in on one side of the card and the city their relatives live in on the other side. A sample of four report cards is selected. Look at whichever cards you wish to test the statement:
[Standard statement] Every person in the sample who lives in London also has a relative who lives in Los Angeles.

[Contraposed statement] Every person in the sample who lives in Seattle also has a relative who lives in Norwich.

2. Drinkers

Only people over the age of eighteen are allowed to drink alcohol in a pub in Britain. A survey is carried out of 100 people in a large public house which identifies their age and whether they are drinking alcohol or a soft drink. Each person's details are put down on a report card with the person's age on one side and their drinking behaviour on the other. A sample of four report cards is selected. To find out if the four people in the sample are obeying the law, look at whichever cards you wish to test the statement:
[Standard statement] Every person in the sample who is drinking alcohol is also over eighteen.
[Contraposed statement] Every person in the sample who is under eighteen is also drinking a soft drink.

3. Rivers

All rivers have either acidic or alkaline water in them. There is a certain chemical which causes fish in acidic rivers to be sick. 100 rivers are investigated and the findings are written down on a report card with whether the river is acidic or alkaline on one side and the health of the fish on the other. A sample of four report cards is selected. To find out if the four rivers in this sample are consistent with their having this type of chemical in the water, look at whichever cards you wish to test the statement:
[Standard statement] Every river in the sample which is acidic also has sick fish.
[Contraposed statement] Every river in the sample which has healthy fish also is alkaline.

4. Objects

[Background] There is a collection of 100 objects each one of which is either 'Grue' or 'Bleen'. Also each object is either 'Smarge' or 'Lall'. Each object is described on a card with 'Grue' or 'Bleen' written on one side and 'Smarge' or 'Lall' written on the other. A sample of four cards is selected.

[Neutral continuation] Look at whichever cards you wish to test the statement:
[Deontic continuation] There is a rule which requires that any object which is 'Grue' must be 'Smarge'. To find out if the four objects in the sample are obeying the rule look at whichever cards you wish to test the statement:
[Causal continuation] There is a group of these objects such that being within this group causes every 'Grue' object also to be 'Smarge'. To find out if the descriptions of the four objects in the sample are consistent with their being in this group look at whichever cards you wish to test the statement:
[Statement] Every object in the sample which is 'Grue' is also 'Smarge'.

5. Diners

[Background] In a restaurant there are 100 people each of whom makes a drinks and a food order. Each person orders either gin or beer and either chips or haddock. Each order is noted down on a card, by the waiter, with the drinks order on one side and the food order on the other side. A sample is taken of orders for food and drink from four people.
[Neutral continuation] Look at whichever cards you wish to test the statement:
[Deontic continuation] The restaurant has a strict rule of etiquette which requires that anyone who eats haddock must drink gin with it. To find out if the four people in the sample are obeying the rule look at whichever cards you wish to test the statement:
[Causal continuation] There is a common eating compulsion which causes anyone who eats haddock to drink gin. To find out if the behaviour of the people in the sample is consistent with their having this compulsion look at whichever cards you wish to test the statement:
[Statement] Every person in the sample who has ordered haddock has also ordered gin.

6. Meters

[Background] In a town there are 100 houses all of which have electricity meters. Each meter is painted either red or violet and may be either inside or outside the house it serves. Inspectors note down the

colour of each meter on one side of a report card and its position on the other side. A sample of four report cards is selected.
[Neutral continuation] Look at whichever cards you wish to test the statement:
[Deontic continuation] A planning law has been passed which requires that any meter on the outside of the house must be red. To find out if the four households in the sample are obeying the law look at whichever cards you wish to test the statement:
[Causal continuation] There is a form of pollution in a certain area which causes every meter on the outside of a house to turn red. To find out if the reports from the four households in the sample are consistent with their living in this area look at whichever cards you wish to test the statement:
[Statement] Every meter in the sample which is outside the house it serves is also red.

NOTES

1. Other forms of confirmation bias have been proposed, for example, the tendency to underweight disconfirming evidence when interpreting ambiguous information: see Klayman (1995) for a discussion of different forms of confirmation bias.
2. In deontic logic, this is equivalent to the *permission* 'If condition q′ is satisfied, then p′ is permitted'. We do not discuss permissions further as they are not used in our experiment.
3. Pointing out that observations of p and q are capable of disconfirming the hypothesis, Klayman and Ha suggest that it is misleading to call this heuristic a 'confirmation bias'; they call it a 'positive test strategy'. We prefer to stick with the established term. The heuristic can be described naturally in terms of thinking what properties an observation would need to have in order to *confirm* the hypothesis *positively* (i.e. the conjunction of p and q), and then sampling cases that have the potential to produce such confirmations. If the heuristic induces the choice of the uninformative q card in Wason's selection task, then in that context it is a *bias*, relative to uncontroversial principles of rationality.
4. We take this to be also the thesis advanced by Oaksford and Chater (1994). Oaksford and Chater claim to explain the selection task evidence in terms of Bayesian rationality, but what they in fact do is to analyse a hypothetical experiment which has a superficial resemblance to the selection task. In this experiment, the p, q, ¬p and ¬q cards of the Wason task are selected at random from a pack of double-sided cards. The subject is required to judge

the truth of the statement ≪If [p], then [q]≫, *as applied to all cards in the pack*. For *this* task, all four cards are informative.

5. As an example of class of stochastic theories which satisfy within-task contraposition neutrality, we offer the following model, which originated in the work of Becker, DeGroot and Marschak (1963), and which has recently been used by Hey and Orme (1994) to test alternative theories of choice under uncertainty. The 'core' of the model is some theory of dominance-respecting preferences over lotteries (e.g. expected utility theory) in which preferences can be represented by a utility function whose domain is the set of lotteries. The individual is assumed to maximize the sum of utility and a random disturbance term, representing error. We can apply this model to our experiment by assuming that the subject reasons about probability in a Bayesian fashion, using the relative frequencies of cards in the pack as priors. The model then implies that if two strategies have the same utility, they are chosen with the same probability.
6. For the first question, the subject is asked to read the first part of a cover story which explains how cards are labelled, is shown a typical four-card layout with one card highlighted, and has to say which two labels might be on the lower face of that card. For the second question, the subject sees a typical statement along with both sides of the four cards in a typical layout; she has to say whether the statement is true or false. The third question asks the subject to say how many points she would have scored after making a particular set of card choices, making a particular judgment, and discovering its truth value. These questions were answered correctly at the first attempt in respectively 92 per cent, 94 per cent and 89 per cent of cases.
7. Cross-task learning might result from analogical reasoning; Gick and Holyoak (1980) give an account of such reasoning.
8. In identifying cards, p always refers to the antecedent in the statement and q to the consequent. Thus, in variable statement tasks, p refers to a different card label in the two treatments. For example, in the case of the Relatives cover story, p refers to ≪London≫ in the standard treatment and to ≪Seattle≫ in the contraposed one.
9. This fact might suggest that a significant minority of our subjects chose {p, ¬q} in every task. In fact, however, only three subjects did this.
10. This significance test, and the others reported in Table 6, should be treated with caution, because the data points do not all derive from different subjects. But the overall pattern in the data is unmistakable.

REFERENCES

Becker, G.M., DeGroot, M.H. and Marschak, J. (1963), Stochastic models of behavior, *Behavioral Science* 8: 41–55.

Börgers, T. (1996), On the relevance of learning and evolution to economic theory, *Economic Journal* 106: 1374–1385.

Börgers, T. and Sarin, R. (1997), Learning through reinforcement and replicator dynamics, *Journal of Economic Theory* 77: 1–14.
Cheng, P.W. and Holyoak, K.J. (1985), Pragmatic reasoning schemas, *Cognitive Psychology* 17: 391–416.
Cheng, P.W. and Holyoak, K.J. (1989), On the natural selection of reasoning theories, *Cognition* 33: 285–313.
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.
Cubitt, R.P. and Sugden, R. (1998), The selection of preferences through imitation, *Review of Economic Studies* 65: 761–771.
Evans, J.St.B.T. (1972), Interpretation and 'matching bias' in a reasoning task, *Quarterly Journal of Experimental Psychology* 24: 193–199.
Evans, J.St.B.T. and Lynch, J.S. (1973), Matching bias in the selection task, *British Journal of Psychology* 64: 391–397.
Fischoff, B. and Beyth-Marom, R. (1983), Hypothesis evaluation from a Bayesian perspective, *Psychological Review* 90: 239–260.
Gick, M.L. and Holyoak, K.J. (1980), Analogical problem solving, *Cognitive Psychology* 12: 306–355.
Griggs, R.A. and Cox, J.R. (1982), The elusive thematic-materials effect in Wason's selection task, *British Journal of Psychology* 73: 407–420.
Hey, J.D. and Orme, C. (1994), Investigating generalizations of expected utility theory using experimental data, *Econometrica* 62: 1291–1326.
Johnson-Laird, P.N., Legrenzi, P. and Legrenzi, M.S. (1972), Reasoning and a sense of reality, *British Journal of Psychology* 63: 395–400.
Klayman, J. (1995), Varieties of confirmation bias, in J. Busemeyer, R. Hastie and D.L. Medin (eds), *Decision Making from a Cognitive Perspective. Psychology of Learning and Motivation* 32: 365–418.
Klayman, J. and Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing, *Psychological Review* 94: 211–228.
Loomes, G. and Sugden, R. (1995), Incorporating a stochastic element into decision theories, *European Economic Review* 39: 641–648.
Manktelow, K.I. and 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 and Over, D.E. (1993), Rationality, utility and deontic reasoning, in K.I. Manktelow and D.E. Over (eds.), *Rationality: Psychological and Philosophical Perspectives*. London: Routledge.
Oaksford, M. and Chater, N. (1994), A rational analysis of the selection task as optimal data selection, *Psychological Review* 101: 608–631.
Reich, S.S. and Ruth, P. (1982), Wason's selection task: verification, falsification and matching, *British Journal of Psychology* 73: 395–405.
Roth, A. and Erev, I. (1995), Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term, *Games and Economic Behavior* 8: 164–212.

Selten, R., Sadrieh, A. and Abbink, K. (1999), Money does not induce risk neutral behavior, but binary lotteries do even worse, *Theory and Decision* 46: 211–249.
Wason, P.C. (1968), Reasoning about a rule, *Quarterly Journal of Experimental Psychology* 20: 273–281
Wason, P.C. and Johnson-Laird, P.N. (1972), *The Psychology of Reasoning: Structure and Content*. Cambridge, Mass.: Harvard University Press.

Addresses for correspondence: Robert Sugden, School of Economic and Social Studies, University of East Anglia, Norwich NR4 7TJ, UK
E-mail: r.sugden@uea.ac.uk