

Focussing in reasoning and decision making

P. Legrenzi^a, V. Girotto^{*,a}, P.N. Johnson-Laird^b

^a*Dipartimento di Psicologia, Università degli Studi di Trieste, Via dell' Università 7, 34123 Trieste, Italy*

^b*Department of Psychology, Princeton University, Green Hall, Princeton, NJ 08544, USA*

Abstract

Our principal hypothesis is that reasoning and decision making are alike in that they both depend on the construction of mental models, and so they should both give rise to similar phenomena. In this paper, we consider one such phenomenon, which we refer to as “focussing”: individuals are likely to restrict their thoughts to what is explicitly represented in their models. We show that focussing occurs in four domains. First, individuals fail to draw inferences in the modus tollens form: if p then q , not- q , therefore not- p , because they focus on their initial models of the conditional, which make explicit only the case where the antecedent (p) and consequent (q) occur. Second, in Wason’s selection task, they similarly tend to select only those cards that are explicitly represented in their initial models of the conditional rule. Third, their requests for information in order to enable them to make a decision about whether or not to carry out a certain action are focussed on the action to the exclusion of alternatives to it. In each of these cases, we show how the focussing bias can be reduced by certain experimental manipulations. Finally, in counterfactual reasoning, focussing underlies individuals’ attempts to imagine an alternative scenario that avoids an unfortunate ending to a story.

Introduction

The classical theory of decision making, whatever its status as a specification of rationality, does not begin to explain the mental processes underlying decisions (e.g., Slovic, 1990). On the one hand, the theory is radically incomplete: it has nothing to say about when one should decide to make a decision, or how one should determine the range of options, or how one should assess the utilities of various outcomes. On the other hand, the theory conflicts with the evidence on

Preparation of this paper was supported by a grant from CNR.

*Corresponding author.

how people reach decisions in daily life: their conspicuous failure to maximize expected utility has led some theorists to worry about human rationality (cf. Lindley, 1985) and others, notably Simon (1959), to argue for a different criterion for human decisions.

An obvious feature of many decisions in daily life is their dependence on reasoning. Consider, for example, buying a new car. The first decision, of course, is to decide that one needs a new car. It is then necessary to determine the range of relevant options, both by trying to specify what one wants and by assessing which cars approximate to these desiderata. Finally, a decision amongst the options must be made. The process is complex, and this summary does not do justice to the adventitious and opportunistic factors that come into play – a particular feature of a car, for example, may not seem desirable until it has been demonstrated to the prospective purchaser. Different cars may have different but conflicting advantages, and it is no easy matter to map them onto a single dimension of utility. A number of component skills, however, can be discerned beneath the variety of approaches that individuals bring to the task of decision making. The key components are as follows:

- Seeking information
- Making hypotheses
- Making inferences
- Weighing advantages and disadvantages
- Applying criteria to make a decision

Hence, at a still deeper level of analysis, various sorts of inferential abilities come into play.

If individuals reason in order to make decisions, then phenomena that occur in reasoning should occur in making judgements and decisions. Logicians distinguish between deductive and inductive reasoning. Human reasoners, however, are seldom interested in deductive validity for its own sake. They aim for conclusions that are true and useful; they often lack sufficient information to reach a valid conclusion, and so are forced to go beyond the information given and to make an induction. In our view, however, the mental machinery underlying both deduction and induction is remarkably similar, and depends not on formal rules of inference but on the construction of mental models (Johnson-Laird, 1993). If so, then the predictions of the model theory should also apply to making decisions. Our aim in this paper is to explore one consequence of this idea – a phenomenon that we refer to as the “focussing effect”, which is an inevitable consequence of the use of models in reasoning.

When individuals construct models, they make explicit as little information as possible in order to minimize the load on working memory. They construct as few explicit models as possible, and they inevitably focus on that information which is

explicit in their models and concomitantly fail to consider other alternatives. Theories based on formal rules cannot make this prediction because they have no machinery for formulating the equivalent of a conclusion based on an initial model of the premises: such conclusions are often invalid, and so they cannot be derived from rules that lead only to valid conclusions. In contrast, the model theory implies that people are inferential *satisficers*: if they come up with a conclusion that fits the available facts, they will tend not to examine other possibilities, with the potentially disastrous consequence of overlooking the correct conclusion. Many of the cognitive errors that have led to real-life disasters have exactly this form. For example, the operators at Three Mile Island explained the high temperature of a relief valve in terms of a leak, and overlooked the possibility that it was stuck open; the master of the English Channel ferry, *The Herald of Free Enterprise*, inferred that the bow doors had been closed, and overlooked the possibility that they had been left open; the engineers at Chernobyl found an erroneous explanation for the initial explosion and overlooked the possibility that the reactor had been destroyed.

The paper begins with an outline of theories of reasoning based on formal rules, and it follows with an account of the contrasting model theory. It then examines focussing in four domains. The first is deduction in the form known as *modus tollens*, e.g.:

If the plane is on course, then the radar should show only water
 The radar is showing a land mass
 Therefore, the plane is not on course

Inferences of this form can be difficult, and indeed the failure to draw this particular conclusion may have contributed to the disaster that befell the ill-fated Flight 007 of Korean Airlines. The cockpit recording shows that crew established the two premises, but instead of drawing the conclusion they changed the topic of conversation (see Overton, 1990). We argue that the difficulty of *modus tollens* is caused in part by focussing, and we report a new experimental study that shows how a simple manipulation helps subjects to defocus. The second domain is a well-known problem in deductive reasoning, Wason's selection task, and we show how focussing leads to errors in its performance. The third domain is decision making. We introduce a new procedure for studying how individuals make decisions. Subjects have to make a simple riskless decision, and their task is to request information, which is provided by the experimenter, until they are able to make the decision. The sequence of questions that they ask reveals the degree of focussing. We report an experiment using this procedure, which shows that the context of decision can play a crucial role in reducing the focussing effect. The fourth domain is counterfactual reasoning, and we report a study that demonstrates the potency of focussing in the construction of counterfactual scenarios.

What is common to all these cases is the strong tendency for individuals to focus on those entities that are explicitly represented in their models (in contravention to the rational requirements of reasoning and decision making), and our attempts to promote the conditions in which alternative possibilities can be made more salient so that focussing disappears or weakens.

FORMAL RULE THEORIES OF REASONING

When people reason deductively, they start with some information – either evidence of the senses or a verbal description – and they assess whether a given conclusion follows validly from this information. In real life there is often no given conclusion, and so they generate a conclusion for themselves. Logic alone is insufficient to characterize intelligent reasoning in this case, because any set of premises yields an infinite number of valid conclusions. Most of them are banal, such as the conjunction of a premise with itself, and no sane individual, apart from a logician, would dream of drawing such a conclusion. Hence, human reasoners are guided by more than logic. The evidence suggests that they tend to maintain the information conveyed by the premises, to re-express it more parsimoniously, and to establish something not directly asserted in a premise. If nothing meets these constraints, they declare that there is no valid conclusion.

Most cognitive scientists assume that deductive reasoning depends on mental rules of inference that are akin to those of the formal method of “natural deduction” (e.g., Braine, 1978; Braine & O’Brien, 1991; Macnamara, 1986; Osherson, 1974–6; Pollock, 1989; Reiter, 1973; Rips, 1983; Sperber & Wilson, 1986). Each connective, such as *if*, *and*, and *or*, has its own formal rules of inference. Deduction accordingly consists in recovering the logical form of premises, and then using the formal rules to try to find a derivation of the conclusion from the premises. If no derivation of the conclusion can be found, then reasoners will respond that the inference is invalid.

For example, given the premises:

If there is a circle then there is a triangle
There is a circle

the valid conclusion:

Therefore, there is a triangle

is easy to deduce, because the mind is equipped with an inferential rule corresponding to *modus ponens*:

If p then q
p
Therefore, q

In contrast, the following premises:

If there is a circle then there is a triangle
There isn't a triangle

often fail to elicit the valid deduction:

Therefore, there isn't a circle

Many intelligent individuals say that nothing follows in this case. The difficulty arises, according to rule theories, because the mind is not equipped with a rule corresponding to this modus tollens inference, and so reasoners must engage in a chain of deductions to arrive at the conclusion. In general, formal rule theories predict that the difficulty of a deduction depends on two factors: the length of the formal derivation, and the unavailability (or difficulty of use) of the relevant rules. Theories differ in their detailed accounts of the difficulty of modus tollens: it introduces negation or an inconsistency, it depends on a reductio ad absurdum, or it depends on some other chain of deductions. There is a consensus, however, on the absence of modus tollens from the mental repertoire of inferential rules (cf. Braine & O'Brien, 1991; Rips, 1983).

THE MENTAL MODEL THEORY OF REASONING

The theory of mental models presents a different view of reasoning (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). The mind does not contain any formal rules of inference akin to a system of "natural deduction". Instead, people reason from their understanding of a situation, and so their starting point is a set of mental models – often, a single model – constructed from perceiving the world or from comprehending discourse. They formulate a conclusion based on the set of models. And they evaluate the conclusion by searching for alternative models that might refute it. We will outline the theory in terms of these stages, and then illustrate them in terms of the modus ponens and modus tollens inferences that we described earlier.

The first stage corresponds, we assume, to the normal processes of perception and comprehension based on the evidence of the senses, discourse, or imagination, and on any relevant general knowledge. The resulting mental models have a structure corresponding, not to the linguistic structure of discourse or to the

syntax of a mental language, but to the structure of the states of affairs, whether perceived, or conceived, that the models represent. A model may be experienced as an image: images are the perceptual correlates of models from a particular point of view. However, many models contain elements, such as negation, that cannot be visualized (see Johnson-Laird & Byrne, 1991, p. 130). What matters is not the subjective experience of the model, which may lie outside conscious awareness, but the structure of the model: sets of entities are represented by sets of mental tokens, the properties of entities are represented by properties of the tokens, and the relations between entities are represented by the relations between the tokens. A model makes *explicit* those objects, properties, and relations that are relevant to potential actions; that is, it makes them available to inference and decision making without the need for further processing. Models, however, make as little as possible explicit because of the limited capacity of working memory. Since their structure corresponds to the structure of situations, models differ from other proposed forms of mental representation, such as semantic networks (e.g., Quillian, 1968) or the propositional representations used for formal rules (e.g., Kintsch, 1974; Braine, 1978). The difference will become clear later when we consider connectives and quantifiers.

The second stage is to formulate a putative conclusion. We assume that it corresponds to the normal process of verbal description. Because conclusions are based on models of the given information, they will not throw semantic information away by adding disjunctive alternatives. The main task is thus to draw a parsimonious conclusion which, if possible, expresses a relation that is not explicit in the premises. If there is no such conclusion, then people say that there is no valid conclusion. Formal theories have neglected the formulation of conclusions, and, like many automated systems of theorem proving, have often assumed that a conclusion is given for evaluation (e.g., Rips, 1983). The model theory, however, has led to a novel algorithm for generating parsimonious conclusions (see Johnson-Laird & Byrne, 1991, p. 183).

The third and critical stage for inference is to evaluate the putative conclusion. The theory assumes that it depends on checking whether alternative models of the situation falsify the conclusion. With simple deductions, human reasoners may be able to anticipate the alternative models, but with complex deductions they can seldom do so, and so they must search for alternative models. Unfortunately, they are not equipped with a comprehensive and systematic search procedure. If there is no alternative model falsifying the conclusion, then it is deductively valid; that is, it must be true given that the premises are true. If there is such a model, then the conclusion is invalid. Prudent reasoners should return to the second stage and try to formulate a new conclusion that satisfies all the models which they have so far constructed. And if they succeed, they should search for further counter-examples, and so on, until they exhaust all possible models (or themselves). If it is uncertain whether there is an alternative model that falsifies a conclusion, then

the conclusion can be drawn on a tentative or probabilistic basis (e.g., Johnson-Laird, 1983; Kahneman & Tversky, 1982). The distinction between deduction and induction is accordingly located primarily in the comprehension and search stages: inductions go beyond the given information because they introduce additional information into models, whereas valid deductions do not; deductions depend on exhaustive searches, whereas inductions do not.

To illustrate how the model theory works, we will reconsider the modus ponens and modus tollens deductions. An initial understanding of a conditional, such as:

If there is a circle then there is a triangle

yields two alternative models. One is an explicit model of the possible state of affairs in which the antecedent holds, and the other is an implicit model in which the antecedent does not hold:

○ △
 . . .

The implicit model, which is denoted by the three dots, is a place-holder that has no initial content and thus minimizes the load on working memory. If necessary, this model may be fleshed out later with an explicit content, but conditional inferences in daily life can often be made from this initial set of models (see Johnson-Laird, Byrne, & Schaeken, 1992). They are consistent with either a one-way conditional, which allows that there may be a triangle without a square, or with a bi-conditional, which makes no such allowance, that is, “if, and only if, there is a circle then there is a triangle”. The one-way conditional is made more explicit in the following models:

[○] △
 . . .

where the square brackets indicate that an element is exhaustively represented; that is, it cannot occur in any model created by fleshing out the implicit model. The consequent is not exhausted, and so it can occur in other models. (Strictly speaking, exhaustion is a relative notion, but we will ignore this qualification here.) The biconditional interpretation calls for both antecedent and consequent to be exhausted:

[○] [△]
 . . .

and in this case reasoners may be more likely to flesh out the implicit model initially (see below). The premise for modus ponens:

There is a circle

eliminates the implicit model from any of these sets, and so the conclusion is immediately forthcoming from the remaining explicit model:

There is a triangle

In contrast, the modus tollens premise:

There is not a triangle

eliminates the explicit model to leave only the implicit model, from which – as many subjects say (e.g., Evans, 1982) – nothing seems to follow. The deduction can be made only by fleshing out the models of the one-way conditional, for example:

○	△
¬○	△
¬○	¬△

or the bi-conditional:

○	△
¬○	¬△

where “¬” represents negation. The premise, “There is not a triangle”, eliminates any model containing a triangle to leave only the model:

¬○	¬△
----	----

This model yields the conclusion:

There is not a circle

which is valid because no model of the premises falsifies it.

The evidence for the model theory of deduction has been presented elsewhere (e.g., Johnson-Laird & Byrne, 1991; Johnson-Laird et al., 1992), and so here we will only summarize it. It shows that the *content* of premises with the same logical form can have a decisive effect on what conclusions people draw. The late Jean Piaget discovered this effect, and introduced a clause in small print – the “horizontal décalage”, essentially a redescription of the phenomenon – to try to sweep it away (e.g., Inhelder & Piaget, 1958). Yet the phenomenon is inimical to formal theories of inference. The evidence also shows that when people reason they are concerned about meaning and truth. They are influenced by what they believe to be true, which affects both the conclusions that they formulate for

themselves and their evaluation of given conclusions. When they draw their own conclusions, they maintain the semantic information from the premises, and treat as improper any conclusions that throw it away by adding disjunctive alternatives. We have carried out experiments on all the main domains of deduction, including inferences based on propositional connectives such as *if* and *or*, inferences based on relations such as *on the right of*, and *in the same place as*, inferences based on quantifiers such as *none*, *any*, and *only*, and meta-logical inferences based on assertions about what is true and what is false. Where the model and rule theories make opposite predictions, the results confirm the model theory and run counter to the formal rule theory. Without exception, all the experiments corroborate the main predictions of the model theory: easy deductions call for one explicit model only; difficult deductions call for more than one explicit model; and erroneous conclusions are usually the result of constructing only one of the possible models of the premises. In the remainder of this paper, we will examine some consequences of focussing on the explicit elements in models.

AN EXPERIMENT ON FOCUSING IN DEDUCTIVE REASONING

Modus ponens can be drawn with a focus on the initial models of the conditional premise; modus tollens calls for the fleshing out of alternative models. Hence, performance of modus tollens should be improved by any manipulation that forces reasoners to cease focussing on the initial models and to make explicit the models where the consequent is false. A recent experiment carried out in collaboration with our colleagues has examined this prediction when subjects have to draw their own conclusions from modus ponens and modus tollens premises (see Girotto, Mazzocco, & Tasso, 1992).

The experimental manipulation was simple. The premises of the inferences were presented either in their traditional order of the conditional premise followed by the categorical premise, or else in an inverted order with the categorical premise followed by the conditional premise. When a conditional premise, such as “If there is a circle at the top of the card then there is a triangle at the bottom of the card”, is presented in the traditional order of modus tollens, then the interpretation of the categorical premise, “There is a square at the bottom of the card”, has to occur while working memory is already pre-occupied with the models of the conditional. Hence, it is difficult for reasoners to flesh out their models of the conditional, and they are likely merely to eliminate the explicit model of the conditional and to replace the implicit model with one representing the categorical premise (“There is a square at the bottom of the card”). It then seems that no conclusion follows.

In contrast, when the categorical premise, “There is a square at the bottom of the card”, occurs first for modus tollens, the subjects represent this contingency

right from the start. When they then begin to interpret the conditional, “If there is a circle at the top of a card then there is a triangle at the bottom of the card”, they can immediately discard the model in which the antecedent and the consequent hold, and thereby free up the processing capacity of working memory. It is now easier to flesh out the implicit model of the conditional to include the explicit combination of the square with the negated antecedent, which yields the conclusion:

There is not a circle at the top of the card

The inverted order with the categorical premise first may help in another way. It may lead individuals to switch round the order of the elements in their models of the conditional, so that the representation of the consequent precedes the representation of the antecedent in working memory. The modus tollens deduction in this case no longer calls for reasoners to work backwards from the most recent item to enter working memory to a previous item: they can work in the same direction as the information entered working memory from the denial of the consequent to the denial of the antecedent (cf. the so-called “figural effect” in syllogistic reasoning as described in Johnson-Laird & Byrne, 1991).

The theory accordingly predicts that performance with modus tollens should be improved in the inverted order of presentation. The effect of the order of premises on modus ponens should be negligible because the inference can be drawn from the initial models of the conditional. If the model theory is correct, then switching round the order of premises should improve modus tollens. However, if formal rule theories are correct, then modus tollens is difficult because it requires a longer and more difficult derivation, and switching round the order of the premises should have no effect on performance. It does not eliminate the negation, or the more complex derivation. It has no effect on logical form.

Method

Design and materials

The experimental design was based on four independent groups of subjects: each group carried out one form of inference (modus ponens or modus tollens) in either the traditional order of premises (conditional followed by categorical) or the inverted order (categorical followed by conditional). The experimental materials had a neutral content designed to insulate them from the subjects’ beliefs or prejudices: they concerned geometrical shapes at the top and at the bottom of a series of cards. A five-page booklet was given to the subjects. The first page stated the general task instructions. The second page described a pack

of cards, each carrying two geometrical shapes: one at the top, and another at the bottom. The page also included an example of a card (with a trapezium at the top and a diamond at the bottom). The following pages presented the problem. The modus tollens version of the problem in the traditional order was as follows (in a translation from the original Italian):

Alberto put some of the cards in a box, on the basis of the following rule:

If there is a circle at the top of the card, then there is a square at the bottom.

Vittorio, who doesn't know what Alberto has done, has taken one of the cards from the box, but he can see only the bottom part, where there is a triangle. Is it possible to draw a conclusion about the upper part (the concealed part) of the card taken from the box by Vittorio? If so, what is the conclusion?

In the inverted order, the problem was the same except that the information about Vittorio's card came first:

Vittorio has taken one of the cards from a box, but he can see only the bottom part, where there is a triangle. Vittorio doesn't know that Alberto put some of the cards in the box, on the basis of the following rule:

If there is a circle at the top of the card, then there is a square at the bottom

Is it possible to draw a conclusion about the upper part (the concealed part) of the card taken by Vittorio? If so, what is the conclusion?

The two conditions presenting the modus ponens problem were similar: in the traditional order, the conditional was stated before the description of Vittorio's card; and in the inverted order, it was given after the description of Vittorio's card. Readers will note that rather than an explicit negation in the modus tollens premises, the materials used a categorical assertion that is inconsistent with the consequent of the conditional. This avoided the problem of counterbalancing the position of the negation. As we will see shortly, it had no critical effect on performance.

Subjects

We tested 92 18–19-year-old students of a high school (in which the major courses concerned scientific matters) of a north Italian town. They were randomly

assigned to one of four conditions: modus ponens in the traditional order ($n = 17$) and in the inverted order ($n = 18$), and modus tollens in the traditional order ($n = 28$) and the inverted order ($n = 29$).

Results and discussion

The percentages of subjects who drew the correct conclusions were as follows:

modus ponens in the traditional order: 88% of subjects
modus ponens in the inverted order: 89% of subjects
modus tollens in the traditional order: 40% of subjects
modus tollens in the inverted order: 69% of subjects

There was a significant improvement in modus tollens when the categorical premise occurred first (rank-sum analysis of variance for contingency tables, $z = 2.1$, $p < .025$), but there was no effect of the order of premises on modus ponens. Hence, it is possible to reduce the focussing effect in the way predicted by the model theory.

Rule theories have no obvious explanation for these results. They could explain them only by introducing some pragmatic principles that would produce different interpretations of the premises as a function of their order of presentation. Subsequently, different inferential rules would be applied to these representations to derive a conclusion (cf. Braine & O'Brien, 1991). Such an explanation, however, is post hoc, whereas the model theory's prediction motivated the experiment in the first place.

Further experiments carried out by Girotto and his colleagues have shown that the same phenomena occur when the modus tollens problems contain categorical premises that are explicitly negative (e.g. "There is not a square"), when the subjects are asked to evaluate given conclusions (e.g., "Is it possible to find a circle on the top, concealed, part of the card?"), and when the inference is presented directly without the context of the two individuals, Alberto and Vittorio. A simple switch in the order of premises can help subjects to defocus, that is, to flesh out their models with explicit information about alternative possibilities.

FOCUSSING IN THE SELECTION TASK

One factor lying behind the success of the model theory's predictions is that reasoners can consider only the information that they have explicitly represented in their models. This factor, we believe, underlies failure in Wason's selection

task. In the original version of the task (Wason, 1966; Wason & Johnson-Laird, 1972), four cards are put in front of a subject, bearing on their uppermost faces a single symbol: A, B, 2, and 3; and the subjects know that every card has a letter on one side and a number on the other side. Their task is to select just those cards that they need to turn over in order to determine whether the following conditional rule is true or false:

If a card has an A on one side then it has a 2 on the other side

The majority of subjects select the A card, or the A and the 2 cards. Surprisingly, they fail to select the card corresponding to the case where the consequent is false: the 3 card. Yet, the combination of an A with a 3 falsifies the rule.

The selection task has generated a large literature, which is not easy to integrate, and one investigator, Evans (1989), has even wondered whether it tells us anything about deduction. He argues that subjects make those selections that merely match the cards mentioned in the rule. Hence, when the rule is negative:

If there is an A then there is *not* a 2

the subjects correctly select the 2 card (which falsifies the consequent). However, deontic conditional rules, such as:

If a person is drinking beer then the person must be over 18

also tend to elicit the correct selection of the cards corresponding to the true antecedent and the false consequent (e.g., Griggs, 1983; Wason, 1983). Several investigators have therefore argued that content-specific rules of inference underlie reasoning in the selection task (e.g., Cheng & Holyoak, 1985).

The model theory explains the selection task in a different way (Johnson-Laird & Byrne, 1991, p. 79):

- (1) The subjects consider only those cards that are explicitly represented in their models of the rule, that is, they focus on these cards.
- (2) They select from these cards those for which the hidden value could have a bearing on the rule, that is, those that are represented exhaustively (with square brackets in our diagrams of models).

Hence, any manipulation that leads to the fleshing out of the models of the conditional with explicit representations of the false consequent will tend to yield insight into the task. An insightful selection may even depend on an explicit

representation of what is impossible given the truth of the conditional, that is, A and not-2.

The conditional, “If there is an A on one side, then there is a 2 on the other side”, yields models of a one-way conditional, or of a bi-conditional, containing only the cards mentioned in the rule:

[A] 2

. . .

or:

[A] [2]

. . .

and so people tend to select the A card, or the A and the 2 card. The model theory is thus compatible with Evans’s matching bias, on the assumption that negation leads to fleshing out the models with the state of affairs that is denied (Wason, 1965). Likewise, experience with the rule about beer drinking helps to flesh out the models with more explicit information:

[drinking beer] over 18

\neg drinking beer [\neg over 18]

. . .

and so subjects will now tend to select the card corresponding to the negated consequent.

In summary, the cause of errors in the selection task is similar to their cause in modus tollens: reasoners focus on what is explicitly represented in their initial models, and so they overlook alternative possibilities. The advantages of this explanation are three-fold. First, it explains why the matching bias disappears with other connectives, such as disjunctions. It does not occur because the initial interpretations of these connectives, unlike those of conditionals, do not contain wholly implicit models. Second, it explains why the matching bias disappears as a result of many different experimental manipulations. Some of these manipulations concern the use of a realistic or deontic content; others have nothing to do with content, such as the use of a linguistically simpler rule, rules concerning single objects, and so on (see Johnson-Laird & Byrne, 1991, pp. 80–81). Third, the explanation’s purview is much wider than that of other theories, which fail to explain the full range of situations in which insight occurs. Such theories include those based on pragmatic reasoning schemas (Cheng & Holyoak, 1985), social contracts (Cosmides, 1989), or “matching” bias as a linguistically determined judgement of relevance (Evans, 1989, p. 33).

Critics may argue: if any manipulation that improves performance is assumed to flesh out mental models, how could the theory ever be disproved? In fact, it is vulnerable to tests in two ways. On the one hand, as we have seen, it predicts that

matching should not occur with connectives that do not initially yield implicit models. On the other hand, it predicts, for example, that the initial representation of a premise of the form, “There is a circle only if there is a triangle”, is more likely to contain a model of the negated consequent:

$$\begin{array}{cc} [\bigcirc] & \Delta \\ \neg\bigcirc & [\neg\Delta] \end{array}$$

Hence, the theory predicts that the “only if” form of the conditional should improve performance in the selection task. So far, we have obtained evidence for this effect only where the content of the conditional concerned deontic matters (see Girotto, Mazzocco, & Cherubini, 1992). What is true is that the model theory does not predict which particular realistic contents will lead to insight into the task, but neither does any other theory. There are likely to be considerable individual differences. Indeed, some individuals perform correctly with abstract materials, and other individuals fail to perform correctly even with deontic materials. A causal explanation of these differences is not in the offing. However, the model theory has the advantage of both generality and parsimony: any manipulation that reduces the subjects’ focus on an initial incomplete representation of the conditional should improve performance. Focussing is a matter of a general theory of reasoning, and it applies to a variety of domains, including the following phenomenon where responses cannot literally match a linguistic constituent of a conditional.

The phenomenon known as “pseudo-diagnosticsity” was demonstrated by Doherty, Mynatt, Tweney, and Schiavo (1979). Their subjects had to determine whether a certain clay pot had come from Coral Island or Shell Island. They were told the features of the particular pot; for example, it had curved handles. They could gain information about the percentages of pots from one island with each of these features or their binary opposites; for example, they could learn that 21% of pots from Coral Island had curved handles, and 79% had straight handles. There were 12 potentially available statistics, but subjects were allowed to select only 6 of them. Few subjects appreciated the need to examine the statistics for both islands. If they formed the hypothesis, say, that the pot came from Coral Island, then they tended to assess how many of the pot’s features were present in pots from Coral Island, and they ignored the critical information about the same features in pots from Shell Island. A similar observation was made by Beyth-Marom and Fischhoff (1983). They presented their subjects with the following scenario: “You have met Mr. Maxwell at a party to which only university professors and business executives were invited. The only thing you know about Mr. Maxwell is that he is a member of the Bear’s Club.” At this point, one group of subjects had to judge the probability that Mr. Maxwell was a university professor, and another group of subjects had to judge whether it was more

probable that he was a university professor than a business executive. The subjects were then asked which of several items of information they would prefer to know in order to make the decision. Most subjects in both groups rated as relevant the information about the proportion of professors at the party who were members of the Bear's Club. In contrast, the proportion of executives at the party who were members of the Bear's Club was judged to be relevant by 78% of the subjects in the second group, but only 54% of the subjects in the first group. The standard interpretation of pseudo-diagnosticity is in terms of attentional bias and a failure to consider alternative possibilities (e.g., Baron, 1988, p. 247). Subjects consider the cases that have both the symptom (membership of the Bear's Club) and the positive category (membership of the set of professors); they ignore the other contingencies (see also Smedslund, 1963; Nisbett and Ross, 1980, p. 92; Evans, 1989, p. 62).

The account of pseudo-diagnosticity in terms of an attentional bias is correct in our view, but we would go further: the underlying mechanism is once again a consequence of models that represent only certain information in an explicit way. A conjecture relating two events, such as a symptom and a disease, corresponds to a conditional. Hence, subjects who are evaluating such a conjecture are likely to represent it by models similar to those of the corresponding conditional:

If ⟨symptom⟩ then ⟨disease⟩

They therefore construct an initial set of models that makes explicit only one model – the model in which the symptom and disease co-occur. Their subsequent search for information is guided by this explicit model. It is the focus of their efforts, and so they concentrate on information about the presence of the symptom and the presence of the disease. They will be prey to the fallacy *post hoc ergo propter hoc*; that is, they will judge that the disease is caused by the symptom because the two co-occur, and they will fail to consider cases where the symptom is present but the disease is absent. Individuals may be able to defocus. For example, the task of judging whether Mr. Maxwell is more likely to be a professor than a business executive calls for a comparison, and so the subjects will construct two contrasting models and consider information about both conditional probabilities to be relevant. The task of defocussing in a diagnostic context accordingly resembles its counterpart in the selection task. The same account in terms of mental models applies to both sets of phenomena.

AN EXPERIMENT ON FOCUSING IN DECISION MAKING

“Utility theory . . . provides at best only part of a normative standard of decision making. The rest of the standard has to do with a thorough search for

alternative actions, goals, and evidence about consequences and their probabilities” (Baron, 1988, p. 289). What focussing implies, however, is that in many circumstances individuals will fail to make a thorough search for alternatives. In particular, if they are faced with the choice of either carrying out a certain action or not carrying it out, then, according to the theory, they will initially construct a model of the action and an alternative model in which it does not occur. The latter will either be implicit or else merely a model in which the action is negated. Hence, individuals will be focussed on the action and search for more information about it in order to reach the decision. They will neglect to search for information about alternative actions, and so contravene the rational standard for decision making. Empirical evidence exists that people fail to consider alternative options and their costs when they are unstated (i.e., people seem to ignore the *opportunity cost principle*, Friedman & Neumann, 1980). In the case of the selection task, there are various manipulations that can help subjects to defocus; for example, the content of the rule can be changed so as to make information about alternatives more available. In the decision task, we can similarly predict that focussing should be reduced by any manipulation that makes alternatives to the action more available.

We tested this prediction in an experiment in which we used a new approach to understanding the mental processes underlying decision making. The subjects’ task was to make a simple riskless decision, for example, whether or not to go to see a certain movie, but they could request any information that they needed in order to make the decision, and the experimenter provided it to them (from one of two pre-established descriptions). The subjects continued to request information until they were able to announce their decision. In the control group, the decision was presented without any background context, and so the subjects should be highly focussed; that is, they should tend to request information about the action and to ignore its alternatives. In the context group, the decision occurred within a particular context: the subjects were asked to imagine that they were visiting, say, Rome for the first time, and that the experimenter was an expert on the city’s tourist attractions. They were then given the task of deciding whether or not to go to a particular event. The context should make alternatives more available, and so the subjects in this group should be less focussed on the action in their requests for information to help them make the decision.

Method

Design

We tested two separate groups of subjects: a control group and a context group, and each subject made five decisions in a different random order about

whether or not to go to a movie, to attend a sporting event, to visit a painting exhibition, to have dinner at a particular restaurant, and to attend an auction.

Materials and procedure

On each trial, the instructions for the control group were of the following form: “You have to decide whether or not (to go to see a certain film). We will give you any information you want until you can make the decision.” The instructions made clear that the decision applied to the subjects’ current action rather than to some indefinite future action. For the context group, they were of the following form:

“Imagine that you are visiting Rome for the first time, but only for one day; and imagine that I have lived in Rome for a long time and have an excellent knowledge of the tourist attractions of the town. Your task is decide whether or not (to go to see a certain film). I will give you any information you want until you can make the decision.”

Each trial of the context group referred to a different well-known tourist city chosen at random from Paris, Athens, Madrid, Tokyo, Rio, and Rome.

For both the control and the context group, we prepared two different descriptions for each event, for example, two types of film (French vs. American, two movie-makers, two actors, two judgements by the critics), two types of restaurant (Chinese vs. Mexican, two ranges of price, two locations), two types of sport event (tennis vs. soccer), and so on. One of the two descriptions was selected at random for each subject. The experimenter used this description as a basic for answering the subjects’ requests for information. Each trial continued until the subject made a decision. Finally, the experimenter asked the subjects to give the reasons why they had made their particular decision. The entire experimental session was tape-recorded, and the main data that we analyzed were the subjects’ questions and final decisions.

Subjects

We tested 16 undergraduates at University College, London, in the experiment, which lasted for about 30 minutes.

Results and discussion

Overall, the subjects asked a mean of 2.55 questions in order to reach a decision. They made a mean of 2.25 positive decisions (out of the five trials) in the

Table 1. *The mean number of questions focussed on the action and the mean number of questions about alternatives for the control group and the context group*

	Questions about the actions	Questions about alternatives
The control group	2.8	0
The context group	2.1	0.25

control group, and a mean of 2.0 positive decisions in the context group. Table 1 presents the mean number of questions focussing on the action and the mean number of questions about alternatives to it for both of the two groups. As the table shows, there is a highly significant focussing effect (which occurs for all subjects, $p = .05^{16}$). However, there was a significant effect of the contextual variable. None of the 8 subjects in the control group ever asked any questions about alternatives to the focussed action, whereas 7 out of the 8 subjects in the context group asked at least one question about alternatives to it (Fisher–Yates exact test, $p < .003$). These questions were typically of the form: “What else is there for me to do in Rome?” Asking about alternatives was correlated with deciding *not* to take the relevant action: the 7 subjects in the context group asked a total of 10 questions about alternatives to the action, but only one of these questions was associated with a positive decision.

When individuals have to choose between going to a certain movie and not going to see it, they tend to focus on the movie in their requests for information to help them to make the decision. We observe the same phenomenon in daily life. There is a natural tendency to focus when a single option is offered for a decision and no obvious alternatives are available. This result is contrary to any theory which assumes that decision makers explicitly consider alternative courses of action. In particular, it is contrary to the classical theory of decision making to which most economists still adhere. Focussing is important to our understanding of how the mind departs from rational principles: if one knows nothing about the alternatives to a particular course of action, one can neither assess their utilities nor compare them with the utility of the action. Hence, one cannot make a rational decision. Of course, *not* taking an action may have consequences (and thus a utility) regardless of the alternative that is actually pursued, but with a few obvious exceptions, such as not taking the antidote to a poison, a negative decision is seldom evaluated as such. The choice between, say, going to the cinema or not going to the cinema is represented by two disjunctive models. The first model is explicit and exhaustive, and so the other model, which corresponds to *not* going to the cinema, can be implicit:

[c]

. . . .

where 'c' denotes a model of going to the cinema. The task for the decision maker therefore appears to be to gather information about c.

The effect of context is to change the nature of the decision, and thus the sort of strategy that subjects are likely to use to make it (Payne, Bettman, Coupey, & Johnson, 1992). Without the tourist context, subjects are likely to compare the movie to some idealized instance. With the tourist context, they are likely to compare it with some other activity. Hence, the two decisions are not psychologically equivalent. The context enables individuals to defocus: it makes alternatives available for fleshing out the implicit model. The subjects can compare the attributes of alternatives. Hence, our results corroborate Payne et al.'s account of the contingent and constructive use of strategies in decision making.

One unexpected finding was the close association between asking about alternatives to the focussed action and reaching a negative decision. One reason may be that subjects who ask for and receive information about alternatives are thereby led to a negative decision. Another reason may be that subjects who are in the process of rejecting a course of action are thereby led to ask about alternatives. There may even be some other underlying factors that bias subjects to ask about alternatives and to reach negative decisions. These potential explanations are not mutually exclusive, but we suspect that the first of them is likely to be the main factor, because subjects reach negative decisions without asking about alternatives, whereas all but one instance of asking about alternatives led to a negative decision. The matter cannot be resolved without further experimentation.

FOCUSSING IN COUNTERFACTUAL REASONING

When people understand discourse, they construct a model of the situation described in the discourse (e.g., Garnham, 1987; Johnson-Laird, 1983). Recent research has shown that events that concern the central character in a narrative are particularly salient: "readers take the character's perspective; they follow the character's thoughts, activating mental images of the same things that the character is thinking about" (Bower & Morrow, 1990, p. 47). Hence, as Morrow, Bower, and Greenspan (1989) have shown, when subjects read the following sentence in a narrative:

John is in the cafeteria and he's thinking of going to the library

they are faster and more accurate in locating items in the library than in locating items in the cafeteria. There seems to be a natural tendency to focus on what is

explicit in a model based on discourse, and particularly those elements relevant to the protagonist's actions.

This phenomenon should have consequences for thinking about counterfactual alternatives to the events in a narrative (e.g., Kahneman & Varey, 1990; Miller, Turnbull, & McFarland, 1990). When readers are forced to envisage such alternatives in order, say, to undo the tragic outcome of a story, they should be focussed on what is most explicit in their models of the story. They should think of alternatives to these events rather than invoking scenarios that are independent of their focus. Consider, for instance, the following synopsis of a story:

A bank employee, Mr. Bianchi, was going home after work, but his progress toward home was delayed by a series of misfortunes (the manoeuvres of a lorry, the passage of a flock of sheep, and a tree trunk lying across the road) and by an intentional decision of his own (to enter a bar to drink a beer). When he arrived home, he found his wife on the floor. He realized that she had suffered a heart attack and that she was dying. He tried to help her, but his efforts were in vain.

What are subjects likely to envisage in order to avoid the story's unfortunate outcome? They might imagine that an ambulance containing a mobile coronary-care unit happened to stop outside Mr. Bianchi's garden at just the right moment, or that his wife called him at work and he was able to arrange for a helicopter to fly her to hospital. These possibilities seem far-fetched because they are not alternatives to any of the explicit events in the narrative. And so subjects should tend not to think of them. They should instead generate counterfactual scenarios that modify explicit elements in their models of the narrative. In particular, the subjects should base their models on the protagonist's actions, and so they should tend to modify precisely these elements, for example, the decision of Mr. Bianchi to enter the bar.

This prediction has been confirmed in an experimental study carried out by Giroto, Legrenzi, and Rizzo (1991). The subjects had to imagine four different alternatives to the story about Mr. Bianchi in which his wife did not die, and to list these alternatives in their order of importance. More than 70% of the subjects' alternatives concerned explicit elements of the story. Each alternative scenario was classified as either active or passive depending on the role played by the protagonist. An active alternative was one in which the protagonist acted or made a decision; for example, "Mr. Bianchi decided not to drink the beer", "Mr. Bianchi helped to remove the tree trunk". A passive alternative was one in which Mr. Bianchi played no such role; for example, "the bar was closed"; "the trunk did not fall". Although modifications could have been of either sort for the bar event, most of the subjects' changes (70%) were of the active sort. For the other events of the story, even those outside the protagonist's control, about 30% of the

changes were of the active sort. Indeed, Mr. Bianchi's decision to enter the bar was the event most often selected for change (25% of all responses) whatever its position in the story and whether it was presented as normal ("Mr. Bianchi, as usual, decided to enter the bar . . .") or abnormal ("Mr. Bianchi, exceptionally for his habits, decided to enter the bar . . ."). Hence, the event that was most available for change was the protagonist's intentional action, regardless of the sequence of the events in the story (*pace* Wells, Taylor, & Turtle, 1987), and of its degree of normalcy (*pace* Kahneman & Miller, 1986). Even when the subjects introduced alternatives from outside the story, these alternatives tended to involve the protagonist.

There is undoubtedly a focus on the protagonist in generating alternative scenarios (see also Kahneman & Miller, 1986), but there is also an additional factor, which may be peculiar to the counterfactual task: the degree to which an event is under the protagonist's control, which may be particularly salient in the model of the story constructed from the protagonist's point of view. It follows that one action by the protagonist will evoke a counterfactual alternative more readily than another action if the former is interpreted as more under the protagonist's control than the latter. This hypothesis was corroborated by Girotto et al. (1991) using a story in which all the episodes were actions performed by the protagonist. An event corresponding to a fully controllable action (entering the bar to drink a beer) was the first event to be changed by 76% of the subjects, in comparison to two other actions forced upon the protagonist by his physical state (deciding to go back to the office for a pair of glasses because his first pair broke, and deciding to stop the car because of an asthma attack).

In sum, as in the case of the construction of a mental model based on a narrative, subjects focus on information relevant to the protagonist when they make inferences about counterfactual states of affairs.

GENERAL DISCUSSION

Focussing is not a new idea in the psychology of thinking. Bruner, Goodnow, and Austin (1956) introduced it in their study of concept attainment: they posited a strategy based on choosing potential instances of an unknown concept by focussing on an initial given positive instance of the concept. This tendency to concentrate on positive instances was also observed in Wason's (1960) studies of hypothesis testing. He gave his subjects an initial triplet of numbers – 2, 4, 6 – and then asked them to find the rule underlying such triplets. They tended to test their hypotheses using positive instances of them. This tendency has been called "confirmation bias", and has been found in inductive reasoning and in other domains such as social judgement (Nisbett & Ross, 1980). Klayman and Ha (1987) reinterpreted the bias as a more general "positive test" strategy, that is, a

tendency to test a hypothesis by checking positive instances of it, which is compatible with the desire to confirm the hypothesis or with the desire to disconfirm it. Evans (1989, p. 42) explains the phenomenon, and the matching-bias data from studies of conditionals, as a consequence of a “positivity” bias. He argues that, unlike Klayman and Ha’s heuristic, this bias is not a deliberate heuristic, but arises from “preconscious heuristic processes which determine the locus of the subjects’ attention”. In our view, all these different heuristics are manifestations of the same underlying mechanism: an inevitable focus on the explicit elements of models.

The focussing mechanism derives from a general theory of reasoning rather than from a hypothesis about some specific experimental results. The general model-based theory is both parsimonious and explicit. However, its explicitness has a cost because it reveals the incompleteness of the theory. One major determinant of what is explicit in mental models is the verbal description of a problem, and another major determinant is the “availability” of knowledge (Tversky & Kahneman, 1973). Yet the theory does not predict the precise circumstances in which an individual will, or will not, flesh out models of a problem explicitly. Hence, it does not explain the circumstances in which individuals are satisfied with a representation of the problem that is only partially explicit. Nevertheless, such partial representations are precisely what lead to the focussing hypothesis, which is a consequence of the construction of models that make explicit only certain items. The hypothesis, in turn, has led us to successful predictions about four different areas: the modus tollens deduction, the selection task, the search for information in riskless decisions, and the construction of counterfactual scenarios. We will review these findings before we consider the general implications of the model theory.

When subjects reason from modus tollens premises of the form *if p then q, not-q*, they often respond that nothing follows from the premises. This error can be explained by a focus on the initial explicit model of the conditional, and a failure to represent the alternative possibilities where the antecedent is false. When the premises are presented in the opposite order, then, as the model theory predicts, the initial representation of the categorical premise immediately eliminates the explicit model of the conditional and thereby helps the subjects to flesh out the implicit model, and so the inference is easier.

In the original form of the selection task, subjects are asked to select those cards that could show whether the following sort of conditional rule is true or false:

If there is an A on one side of a card, then there is a 2 on the other side

They tend to select either the A card alone, or the A and the 2 card. The focussing hypothesis explains the phenomenon in terms of what is made explicit in

the models of the conditional. Hence, any manipulations that lead subjects to flesh out these models should enhance performance, and indeed a variety of such manipulations has been found to produce this effect.

A decision is sometimes defined as a conscious choice between at least two possible courses of action (e.g., Castles, Murray, & Potter, 1971, cited by Fox, 1990). Paradoxically, when subjects have to decide whether or not to carry out a certain action, the effect of focussing is to reduce the decision to a choice between one and a half alternatives. The models of the decision contain explicit information about the action, but not necessarily any explicit information about *not* taking the action. The subjects are thus focussed on the action itself, and they seek information only about its attributes in order to help them to make the decision. Subjects are able to defocus, however, when the scenario for the decision makes alternatives to the action more available. They seek information about these alternatives in accordance with the normative standards for decision making.

When subjects have to envisage an alternative counterfactual scenario that avoids an unfortunate outcome, they are once again focussed on the events explicitly represented in their models of the narrative. Such models, however, tend to represent events from the point of view of the protagonist: the protagonist's actions are particularly salient, and especially those that are likely to be represented as under his or her control. The extent to which this latter factor is local to the particular counterfactual task calls for further empirical investigation.

Focussing should lead to predictable requests for certain sorts of information in making decisions between two explicit alternatives. For example, suppose individuals have to choose between two alternative vacation resorts:

Resort A has good beaches, plenty of sunshine, and is easy to get to
Resort B has good beaches, cheap food, and comfortable hotels

What further information are individuals likely to request in order to choose between the two resorts? The focussing hypothesis implies that they will seek to build models that flesh out the missing attributes. They know, for example, that resort A has plenty of sunshine, but they know nothing about the weather at resort B, and so they will seek information about this attribute. We can further predict that once the specified attributes have been fleshed out in this way, the subjects will be able to make a decision provided that one resort has all the positive attributes of the other, and some additional positive attributes. In summary, the initial specification of the decision acts as a focus for both the information that subjects seek and their ultimate decision, and consequently they will tend to overlook other attributes, such as hostility to tourists, a dictatorial government, or rampant food poisoning, that are not included in the original specification. Yet, these attributes might well influence their decisions in other

instances. We note that one factor in the early success of Ross Perot in his bid for the US Presidency appears to have been the attractiveness of those attributes that he revealed in his television appearances. Those who supported him appeared to have focussed on these attributes, and to have neglected those others – such as most matters of policy – that he chose not to reveal.

Focussing may well account for a number of so-called “framing” effects in decision making and their analogues in deductive reasoning. The same decision can be framed in different ways that will lead individuals to build different models of the choices. In a study by Tversky and Kahneman (1981), a choice between two different programs for combatting a disease was framed in terms of lives saved:

Program A will save 200 lives (out of 600)

Program B has a $1/3$ probability of saving 600 lives and a $2/3$ probability of saving no one

and the majority of subjects preferred program A. However, when the same choice was framed in terms of numbers who die:

Program A will lead to 400 deaths (out of 600)

Program B has a $1/3$ probability of no one dying and a $2/3$ probability of 600 deaths

the majority of subjects preferred program B. Tversky and Kahneman explain this phenomenon in terms of the respective reference points of the two ways of framing the problem, and the shape of the function relating gains and losses to subjective value: a loss hurts more than a gain pleases. The reference point for the first problem is 600 lives lost, and so the choices lie on the concave part of the function for gains, whereas the reference point for the second problem is 0 lives lost, and so the choices lie on the convex part of the function for losses. It may be, however, that part of the difference is merely in the models that subjects build for the respective versions of option A: in the first problem, a model that makes explicit 200 lives saved, and, in the second problem, a model that makes explicit 400 deaths. A is accordingly attractive in the first problem, but unattractive in the second problem.

The framing of problems in deductive reasoning similarly influences performance. According to the model theory, the assertion:

All the authors are boxers

is represented by a single model of the following sort:

[a]	b
[a]	b
.	.
.	.

where each line in this diagram represents a different individual in the same model, an “a” represents an author, and a “b” represents a boxer. The square brackets indicate that a set has been exhaustively represented in the model; that is, members of the set of authors cannot occur elsewhere in the model; and the three dots allow for other sorts of individual yet to be made explicit. The number of individuals remains arbitrary, but it is likely to be small. The assertion:

Only the boxers are authors

has the same truth conditions as the previous assertion, but, according to the theory, it makes explicit right from the start that anyone who is *not* a boxer is also *not* an author. Hence, its initial model is of the following sort:

b	[a]
b	[a]
[¬b]	¬a
.	.
.	.

and so the models of the two assertions are equivalent in content, though the equivalence is not immediately apparent. Deductions based on what is explicit in a model should be easier than those that depend on fleshing out implicit information. It follows that the premises:

All athletes are boxers
Mark is an athlete

should readily yield the conclusion:

Mark is a boxer

whereas the premises:

All athletes are boxers
Mark is not a boxer

should less readily yield the conclusion:

Mark is not an athlete

For the corresponding problems based on “only”, there should be no difference between the premises:

Only boxers are athletes
Mark is an athlete

and:

Only boxers are athletes
Mark is not a boxer

because the models contain explicit negative information right from the start. The results from an experiment corroborated this hypothesis (see Johnson-Laird & Byrne, 1989). Hence, the framing of a problem can lead to models that differ in what they make explicit about the situation, and this difference in turn can affect the conclusions that reasoners draw.

When individuals make a choice between two alternatives, then choosing one alternative is logically equivalent to rejecting the other. Yet, as Shafir (1991) has shown, the two tasks of choosing and rejecting may not be psychologically equivalent. When subjects have to select an option, they focus on the positive attributes of the alternatives; but when they have to reject an option, they focus on the negative attributes of the alternatives. Shafir was accordingly able to offer choices where one and the same alternative had both many positive attributes and many negative attributes in comparison to the other alternative. The result was that subjects tended both to choose this alternative in the choice task and to reject it in the rejection task. We can predict that there should be a similar effect on seeking and evaluating information about the options. In the context of choosing an option, subjects will pay more attention to positive attributes, whereas in the context of rejecting an option, they will pay more attention to negative attributes.

The application of the model theory to the study of decision making should permit one to make predictions on the basis of what is known about the use of mental models in deductive reasoning, and so we can predict that model-based phenomena other than those of focussing should occur in making decisions. In this final section of the paper we will consider the possibility. The single most robust phenomenon in deductive reasoning is that deductions that call for more than one model are harder than those that call for only one model. We have observed this phenomenon in studies of propositional, relational, syllogistic, and multiply quantified reasoning (for a review, see Johnson-Laird & Byrne, 1991). Deductive performance can be tested to the point where it breaks down merely by increasing the number of disjunctive models that have to be constructed (see the study of “double disjunctions” in Johnson-Laird et al., 1992). We can expect that decisions will likewise grow more difficult as a function of the number of options.

Indeed, like deduction, there can be a breakdown in rationality as soon as there are two explicit alternatives to choose from – as shown by the so-called “disjunctive effect” in decision making (see Tversky & Shafir, 1992, and the paper in this issue by Shafir, Simonson, & Tversky). An everyday example of this phenomenon occurred during the early stages of the recent US Presidential campaign: at one point, the opinion polls revealed that Bush would lose to a Democrat, but that he would beat each individual Democratic candidate. In general, as Shafir and Tversky (1992) have shown experimentally, subjects may choose a particular option when a certain event occurs, and when it does not occur. Yet, when the outcome is unknown, they do not choose it. In our terms, the need to hold in mind the disjunctive alternatives makes the inferential task difficult: subjects are unable or reluctant to derive the consequences of each alternative and to assess what they have in common. Part of this difficulty, as Shafir and Tversky argue, may be that the reasons for the choice are quite different for the alternative outcomes, for example, choosing a Hawaiian vacation as a reward for passing an exam as opposed to as a consolation for failing the exam. In attempting to make the decision when the outcome is unknown, there is no simple reason for choosing the vacation. The disjunctive alternatives appear to conflict.

We can predict at least one other disjunctive effect in decision making: if the information available about a particular option is disjunctive in form, then the resulting conflict or load on working memory will make it harder to infer a reason for choosing this option in comparison to an option for which categorical information is available. The harder it is to infer a reason for a choice, the less attractive that choice is likely to be.

CONCLUSION

Focussing is a widespread phenomenon in reasoning and decision making. Certain experimental manipulations, however, enable subjects to defocus: they flesh out their models of alternative possibilities, and use these models in making deductions and decisions. Neither of these effects – focussing and defocussing – is predicted by the respective classical theories of the two domains, that is, theories of reasoning based on formal rules of inference, and theories of decision making based on the maximization of expected utilities. Both phenomena, however, are predicted by the theory that reasoning depends on the construction of mental models or scenarios of the relevant situations.

References

- Baron, J. (1988). *Thinking and deciding*. Cambridge, UK: Cambridge University Press.
 Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology*, 45, 1185–1197.

- Bower, G.H., & Morrow, D.G. (1990). Mental models in narrative comprehension. *Science*, 247, 44–48.
- Braine, M.D.S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, 85, 1–21.
- Braine, M.D.S., & O'Brien, D.P.O. (1991). A theory of "if": A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, 98, 182–203.
- Bruner, J.S., Goodnow, G., & Austin, G. (1956). *A study of thinking*. New York: Wiley.
- Castles, F.G., Murray, D.I., & Potter, D.C. (Eds.) (1971). *Decisions, organizations and society*. Harmondsworth, UK: Penguin.
- Cheng, P.W., & Holyoak, K.J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391–416.
- 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.
- Doherty, M.E., Mynatt, C.R., Tweney, R.D., & Schiavo, M.D. (1979). Pseudodiagnosticity. *Acta Psychologica*, 43, 11–21.
- Evans, J.St.B.T. (1982). *The psychology of deductive reasoning*. London: Routledge & Kegan Paul.
- Evans, J.St.B.T. (1989). *Bias in human reasoning*. Hillsdale, NJ: Erlbaum.
- Fox, J. (1990). Automating assistance for safety critical decisions. In D.E. Broadbent, A. Baddeley, & J.T. Reason (Eds.), *Human factors in hazardous situations* (pp. 107–119). Oxford: Clarendon Press. (*Philosophical Transactions of the Royal Society of London B*, 327, 555–567.)
- Friedman, L.A., & Neumann, B.R. (1980). The effects of opportunity costs on project investment decisions: A replication and extension. *Journal of Accounting Research*, 18, 407–419.
- Garnham, A. (1987). *Mental models as representations of discourse and text*. Chichester: Ellis Horwood.
- Giroto, V., Legrenzi, P., & Rizzo, A. (1991). Event controllability in counterfactual thinking. *Acta Psychologica*, 78, 111–133.
- Giroto, V., Mazzocco, A., & Cherubini, P. (1992). Judgements of deontic relevance in reasoning: A reply to Jackson and Griggs. *Quarterly Journal of Experimental Psychology*, 45A, 547–574.
- Giroto, V., Mazzocco, A., & Tasso, A. (1992). The effect of premise order effect on conditional reasoning: A test of the mental model theory. Unpublished manuscript, Department of Psychology, University of Trieste.
- Griggs, R.A. (1983). The role of problem content in the selection task and in the THOG problem. In J.St.B.T. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge & Kegan Paul.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. London: Routledge & Kegan Paul.
- Johnson-Laird, P.N. (1983). *Mental models*. Cambridge, UK: Cambridge University Press.
- Johnson-Laird, P.N. (1993). *Human and machine thinking*. Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P.N., & Byrne, R.M.J. (1989). Only reasoning. *Journal of Memory and Language*, 28, 313–330.
- Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P.N., Byrne, R.M.J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, 99, 418–439.
- Kahneman, D., & Miller, D.T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93, 136–153.
- Kahneman, D., & Tversky, A. (1982) The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). Cambridge, UK: Cambridge University Press.
- Kahneman, D., & Varey, C.A. (1990). Propensities and counterfactuals: The loser that almost won. *Journal of Personality and Social Psychology*, 59, 1101–1110.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.
- Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, 94, 211–228.
- Lindley, D.V. (1985). *Making decisions*, 2nd edn. Chichester: Wiley.
- Macnamara, J. (1986). *A border dispute: The place of logic in psychology*. Cambridge, MA: MIT Press.

- Miller, D.T., Turnbull, W., & McFarland, C. (1990). Counterfactual thinking and social perception: Thinking about what might have been. In P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 305–331). Orlando, FL: Academic Press.
- Morrow, D.G., Bower, G.H., & Greenspan, S.L. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language*, 28, 292–312.
- Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgement*. Englewood Cliffs, NJ: Prentice-Hall.
- Osherson, D.N. (1974–6). *Logical abilities in children* (Vols. 1–4). Hillsdale, NJ: Erlbaum.
- Overton, W.F. (1990). Competence and procedures: Constraints on the development of logical reasoning. In W.F. Overton (Ed.), *Reasoning, necessity and logic: Developmental perspectives*. Hillsdale, NJ: Erlbaum.
- Payne, J.W., Bettman, J.R., Coupey, E., & Johnson, E.J. (1992). A constructive process view of decision making: Multiple strategies in judgment and choice. *Acta Psychologica*, 80, 107–141.
- Pollock, J. (1989). *How to build a person: A prolegomenon*. Cambridge, MA: MIT/Bradford Books.
- Quillian, M.R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing*. Cambridge, MA: MIT Press.
- Reiter, R. (1973). A semantically guided deductive system for automatic theorem-proving. *Proceedings of the Third International Joint Conference on Artificial Intelligence*, 41–46.
- Rips, L.J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, 90, 38–71.
- Shafir, E. (1991). Choosing versus rejecting: Why some options are both better and worse than others. Unpublished manuscript, Department of Psychology, Princeton University.
- Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology*, 24, 449–474.
- Simon, H.A. (1959). Theories of decision making in economics and behavioral science. *American Economic Review*, 49, 253–283.
- Slovic, P. (1990). Choice. In D.N. Osherson, & E.E. Smith (Eds.), *An invitation to cognitive science. Vol. 3: Thinking* (pp. 89–116). Cambridge, MA.: MIT Press.
- Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, 4, 165–173.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Basil Blackwell.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207–232.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453–458.
- Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science*, 3, 305–309.
- 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. (1965). The context of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, 4, 7–11.
- Wason, P.C. (1966). Reasoning. In B.M. Foss (Ed.), *New horizons in psychology*. Harmondsworth, UK: Penguin.
- Wason, P.C. (1983). Realism and rationality in the selection task. In J.St.B.T. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge & Kegan Paul.
- Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford; Cambridge, MA: Harvard University Press.
- Wells, G.L., Taylor, B.R., & Turtle, J.W. (1987). The undoing of scenarios. *Journal of Personality and Social Psychology*, 53, 421–430.