Discussion

Modal reasoning, models, and Manktelow and Over

Philip N. Johnson-Laird
Princeton University, Princeton, NJ 08544, USA

Ruth M.J. Byrne
Department of Psychology, Trinity College, University of Dublin, Dublin 2, Ireland

Abstract

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

Manktelow and Over (1991) argue that their studies of Wason's selection task favor explanations of deontic reasoning based on mental models, but that such theories need to incorporate utilities. This theoretical note proposes a simpler explanation of the phenomena: subjects in the selection task consider only those cards that are explicitly represented in their models of the conditional, and so insight into the task depends on constructing fully explicit models. Such models for modal conditionals of the form, If p occurs then q may occur *are:*

$$
\begin{array}{cc}
p & q \\
p & \neg q \\
\neg p & \neg q
\end{array}
$$

Each line denotes a separate model, and the models represent either what is possible, or, in the deontic interpretation, what is permissible. A deontic rule is accordingly violated by the contingency: $\neg p$ and q, for example the rule, "If you spend more than £100, then you may take a free gift" is violated by taking the free gift (q) but not spending more than £100 ($\neg p$). If the rule is interpreted as a

Correspondence to: Dr. P.N. Johnson-Laird, Department of Psychology, Princeton University, Green Hall, Princeton, NJ 08544, USA. Tel: (609) 258 4432; Email: phil@clarity.princeton.edu.

bi-conditional, then the second of the models, p and ¬q, is also now a violation, for example spending more than £100 (p) but not getting the free gift (¬q). Manktelow and Over's instructions lead subjects to focus on one or other of the two sorts of violations of the rule. There is accordingly no need to introduce utilities into models in order to explain the phenomena.

Introduction

Recent studies of choice have shown that subjects choose among alternatives, not by considering their expected utilities – the *sine qua non* of classical theories of decision making – but by constructing reasons to justify one choice as opposed to others (see Shafir, 1991; Tversky and Shafir, 1991). Ironically, just as students of choice are abandoning utilities for reasons, students of deductive inference are introducing utilities to supplement reasons. In particular, Manktelow and Over (1991) argue that subjects engaged in deontic reasoning attach "different utilities to the outcomes of relevant outcomes they or others might perform". In this note, our aim is to argue for an alternative explanation, which we believe is in line with Manktelow and Over's account, but which provides a simpler explanation of the phenomena. In fact, this revision avoids the need to introduce utilities.

The deontic selection task: Manktelow and Over's results

Manktelow and Over report three experiments that use Wason's (1966) selection task to investigate deontic reasoning, that is, reasoning about obligations and permissions. They gave subjects assertions that express conditional permissions, such as:

If you spend more than £100, then you may take a free gift.

The subjects had in front of them four cards that each represented a receipt and whether or not the customer had taken the free gift:

The shopper spent more than £100   (p)
The shopper did not spend more than £100   (¬p)
The shopper took the free gift   (q)
The shopper did not take the free gift   (¬q)

We have abbreviated the cards in terms of p, q, and their negations, assuming that the conditional has the form: if p then q. The task was to select those cards which, if turned over, might show that the shop had violated the rule, or, in another condition, that the shopper had violated it. When the subjects chose

cards to test whether the shop had broken the rule, they tended to select the p and ¬q cards:

spent more than £100; did not take the free gift

The subjects also made this selection when they tested whether the shoppers had not taken what they were entitled to. However, when the subjects tested whether the shopper had *broken* the rule, they tended to select the ¬p and q cards:

did not pay more than £100; took the free gift

They also made this selection when they tested whether the store had been giving customers more than they were entitled to.

The content of a conditional affects performance on the selection task. In testing the truth or falsity of an abstract conditional, such as "If there is an A on one side of the card, then there is a 2 on the other side of the card", subjects seldom grasp the need to select the card representing the false consequent (the ¬q card). In contrast, they do tend to select this card when the materials have a sensible everyday content (see, for example, Evans, 1989; Griggs & Cox, 1982; Wason, 1983; Wason & Shapiro, 1971). This difference in performance, as Manktelow and Over (1991) have argued, is difficult to account for in terms of a theory based on formal rules of inference. A more plausible explanation, which was advanced by Cheng and Holyoak (1985), is that whenever the conditional engages certain pragmatic schemas for reasoning, especially deontic ones, subjects are likely to grasp the need to select the ¬q card (see also Cheng, Holyoak, Nisbett, & Oliver, 1986). Cosmides (1989) has gone further and argued that human beings are equipped with an innate procedure for detecting cheaters, that is, individuals who violate the principle that if you take a benefit, then you pay a cost.

Manktelow and Over draw three principal conclusions. First, subjects naturally interpret many realistic conditionals as deontic. In the selection task, they then test, not whether the conditional is true or false, but whether it has been violated by failures to conform. Psychologists should therefore, they say, "drop the narrow idea of a unique logically correct response . . . and speak more broadly instead of various *rational* responses for different contents and contexts" (Manktelow & Over, 1991, p. 101). Second, deontic reasoning cannot be properly explained by Cheng and Holyoak's schemas, because these schemas themselves contain such unanalyzed deontic terms as "may" and "must" (see also Jackson & Griggs, 1989; Johnson-Laird & Byrne, 1991, p. 79). Likewise, Cosmides' theory is too narrow, because subjects make insightful selections that do not depend on a "check for cheaters" procedure, for example when the subjects tested whether the shoppers had failed to take the free gift to which they were entitled. Third, the results favor an explanation in terms of mental models (Johnson-Laird, 1983), but this theory, even in its most recent formulation (Johnson-Laird & Byrne, 1991), overlooks the

importance of utilities, subjective probabilities, and the variety of pragmatic effects on deontic reasoning.

A model theory of Wason's selection task

Although we agree with Manktelow and Over about the deontic interpretation of conditionals, and with the need to explain the phenomena in terms of mental models, we do not believe that their results necessarily call for the introduction of subjective expected utilities. Hence, we will re-examine the phenomena in the light of the model theory of deontic reasoning (Johnson-Laird, 1978) and of the selection task (Johnson-Laird & Byrne, 1991). According to the model theory, a simple indicative conditional, such as:

If there is a letter "A", then there is a number "2"

is normally represented by one explicit model satisfying the antecedent and consequent, and one implicit model that merely allows for the case in which the antecedent is not satisfied:

[A]   2
    . . .

The square brackets indicate that the A has been exhaustively represented in relation to occurrences of 2, that is, A cannot occur in any other model of the conditional (see Johnson-Laird, Byrne, & Schaeken, 1991). Subjects in the selection task, the theory assumes, consider only those cards that are explicitly represented in their models of the conditional, and so they consider only the A and 2 cards. They will select the A card whether they wish to verify or to falsify the conditional, and they will select the 2 card if they wish to verify the conditional. Indeed, most subjects select the A and 2 cards, or the A card alone (Wason, 1966, 1983).

The models of a conditional can be fleshed out explicitly as three distinct models, which each represent a different possibility:

A   2
¬A   2
¬A   ¬2

The subjects in the selection task will now consider the ¬A and the ¬2 cards. The ¬A card can occur with either 2 or ¬2 and so it cannot reveal that the rule is true or false: subjects will be unlikely to add it to their selections made on the basis of the initial models. The ¬2 card would falsify the rule if it occurred with the A card, and so subjects should now be likely to add it to their selections.

The model theory predicts that any experimental manipulation that leads

subjects to flesh out their models explicitly should also lead them to select the card representing the negated consequent of the conditional (¬q). One such manipulation is to introduce a familiar deontic framework (Johnson-Laird, Legrenzi, & Legrenzi, 1972). This framework brings to mind instances of possible violations, and so the subjects are likely to flesh out their models explicitly with such instances. There are, however, other manipulations that also lead to fleshing out models explicitly but that have nothing to do with deontic matters, for example a simpler conditional in the selection task (Wason & Green, 1984), or simpler descriptions in testing the accuracy of a description of the contents of an envelope – a task that is logically isomorphic to the selection task (Oakhill & Johnson-Laird, 1985). Our first major conclusion is accordingly that no account of the selection task based solely on deontic considerations can be wholly correct (*pace* Cheng & Holyoak, 1985; Cosmides, 1989).

A model theory of modal reasoning

A modal assertion such as:

John may leave

is open to at least two different interpretations. It has an *epistemic* interpretation that refers to the possibility that John will leave, and it has a *deontic* interpretation that refers to his permission to leave. Johnson-Laird (1978) therefore argued that modals are interpreted with respect to two different frameworks: an epistemic framework concerning what is possible and a deontic framework concerning what is permissible.

In the epistemic domain, it is necessary to distinguish four sorts of situations: facts, which correspond to the actual world; possible situations, which may or may not occur; counterfactual situations, which once were possible but did not occur; and impossible situations, which given the facts could not have occurred. It is necessary to represent the status of a particular mental model in relation to these four distinct alternatives. Modal logics notoriously fail to square with everyday language, for example "I am laughing but possibly I am not laughing" is acceptable in modal logic, but anomalous in everyday language. The heart of the discrepancy is that modal logics do not distinguish the four sorts of situation. The acceptable assertion "I am laughing but I might not have been laughing" couples a factual assertion with a counterfactual, not with a possibility. The models of the conditionals in the previous section all represent possibilities, but the normal interpretation of a counterfactual conditional, such as:

If there had been a letter "A", then there would have been a number "2"

calls for the following initial models:

Fact: ¬A   ¬2
Counterfactual situations: [A]   2
                                   . . .

where a model of what is impossible given the truth of the counterfactual is:

Impossible situation: A   ¬2

The distinctions in epistemic status are independent of the contrast between actuality and fiction: fictional discourse also calls for exactly the same four alternatives relative to the world that is depicted in the fiction.

A conditional with an epistemic consequent, such as:

If there is letter "A", then there may be a number "2"

calls for a different set of models from those of the corresponding indicative conditional lacking the modal consequent. In particular, the epistemic consequent calls for the models:

Possible situations: A   [2]
                                . . .

They can be fleshed out explicitly as:

Possible situations: A   2
                          A   ¬2
                         ¬A   ¬2

The contingency in which there is ¬A and 2 is impossible on pain of tautology, that is, if all four models were possible then nothing could falsify the conditional and so subjects would not choose any cards in the selection task. If the subjects build the three explicit models above, then they should choose the ¬A card and the 2 card in the selection task, that is, the cards corresponding to the false antecedent (¬p) and the true consequent (q). This prediction has been corroborated with deontic conditionals, but has yet to be tested with epistemic ones (Cosmides, 1989; Manktelow & Over, 1991).

Deontic assertions concern what is permissible and impermissible, which are both normally possible. Thus, the conditional:

If you spend more than £100, then you may take a free gift

has the modal auxiliary, "may", in its consequent, and calls for the same models as the epistemic conditional. Deontic conditionals are often sufficiently familiar to elicit fully explicit models:

Permissible situations: £100   gift
                             £100   ¬gift
                            ¬£100   ¬gift

where "£100" represents a model of the antecedent event (you spend more than £100) and "gift" represents a model of the consequent event (you take the free gift). What is impermissible because it violates the conditional is to take the gift without having spent more than £100:

Impermissible situation: ¬£100 gift

A stronger interpretation of the rule treats it as a bi-conditional, with the resulting models:

Permissible situations: £100 gift
                         ¬£100 ¬gift

What is now impermissible is either of the following two situations:

Impermissible situations: £100 ¬gift
                          ¬£100 gift

The first case could arise because you declined to take the free gift even though you had spent more than £100, or, alternatively, because the store welched on its rule. Hence, if subjects in the selection task are asked to choose those cards that would allow them to check for these possibilities, then in both cases they should select the £100 card and the ¬gift card, that is, the true antecedent and the false consequent. This pattern is precisely the one observed by Manktelow and Over. The second of the two impermissible cases above could arise because you took the free gift even though you were not entitled to it, or because the store allowed you to take it even though you were not entitled to it. Hence, if subjects are asked to choose those cards that would allow them to check for these possibilities, then in both cases they should select the ¬£100 card and the gift card, that is, the false antecedent and the true consequent. This pattern is also precisely the one observed by Manktelow and Over. Our second major conclusion is that epistemic and deontic conditionals can be represented by models, and the instructions in the selection task can lead subjects to focus on one set of models rather than another.

An unexpected result

In their first experiment, Manktelow and Over obtained an unexpected result that our theory explains. The conditional permission was as follows:

If you tidy your room, then you may go out to play

which can be interpreted according to the following models:

Permissible situations: tidy play
                         ¬tidy ¬play

What is impermissible is either of the following two situations:

Impermissible situations:	tidy	¬play
	¬tidy	play

In one condition, the subjects were asked to select those cards that would show whether the mother who had laid down the rule had broken it. She could have brought about either of the impermissible situations: the first if she refused to allow the child to play even though the child has tidied the room, and the second if she allowed the child to play even though the child has not tidied the room. The second violation is, of course, less serious: it is a case of lenience that the child should not object to. Hence, subjects are likely to select either those cards that would reveal the first and more serious violation or all four cards that would reveal both sorts of violation. This pattern of selections is exactly what Manktelow and Over observed in the experiment. They attributed the selection of all four cards to the fact that the scenario concerned only the mother and child. They therefore introduced a "third party" into their next experiment, and observed only the pattern of selections corresponding to the serious violation: the tidy card and the ¬play card. What they have overlooked, however, is that a change in the instructions now makes clear that only the serious violation is at stake: "Your brother has been getting angry: he thinks that sometimes Mother has been unfair to him and not kept her side of the rule while he has kept his." Our third major conclusion is that there is no need to introduce a third party in order to eliminate the unexpected result: the phenomena depend solely on whether the instructions concern the agent merely breaking the rule (select all four cards) or the agent acting unfairly (select two cards).

Conclusions

We have argued that performance in the selection task can be improved without introducing a deontic framework. However, when the modal auxiliary, "may", occurs in the consequent of a conditional of the form:

If p occurs then q may occur

then, given a plausible everyday content, reasoners are likely to construct a fully explicit set of models of the form:

Possible situations:	p	q
	p	¬q
	¬p	¬q

and to select the ¬p and the q cards to test the truth of the conditional. The same argument applies to deontic conditionals, but now in many cases general knowl-

edge leads to the bi-conditional interpretation:

Permissible situations: p   q
                           ¬p   ¬q

Manktelow and Over asked subjects to focus on one or other of the two sorts of possible violations of the rule, and the subjects made their selections accordingly. This explanation accounts for all of the results. It is consistent with the underlying tenor of Manktelow and Over's arguments, but it has the advantage of parsimony. There is no need to abandon the notion of a correct rational solution to the selection task, and there is no need to introduce utilities into models. Utilities may, of course, be needed for other reasons, but we suspect that the case for abandoning them and other forms of *given* valuations of situations is likely to become still stronger. Human reasoners choose between real-life alternatives much as they make choices in the selection task: they *construct* models based on general and specific knowledge.

References

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

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

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.

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

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

Jackson, S.L., & Griggs, R.A. (1989). The elusive pragmatic reasoning schemas effect. *Quarterly Journal of Experimental Psychology*. *42A*, 353–373.

Johnson-Laird, P.N. (1978). The meaning of modality. *Cognitive Science*, *2*, 17–26.

Johnson-Laird, P.N. (1983). *Mental models*. Cambridge, MA: Harvard University Press; Cambridge, UK: Cambridge University Press.

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

Johnson-Laird, P.N., Byrne, R.M.J., & Schaeken, W. (1991). Propositional reasoning by model. *Psychological Review*, in press.

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

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

Oakhill, J.V., & Johnson-Laird, P.N. (1985). Rationality, memory, and the search for counterexamples. *Cognition*, *20*, 79–94.

Shafir, E. (1991). Choosing versus rejecting: Why some options are both better and worse than others. Unpublished MS, Department of Psychology, Princeton University.

Tversky, A., & Shafir, E. (1991). The disjunction effect under uncertainty. Unpublished MS. Department of Psychology, Stanford University.

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

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