Thinking & Reasoning
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ptar20

A model point of view

P. N. Johnson-laird [a] & Ruth M.J. Byrne [b]
[a] Princeton University, USA
[b] University of Dublin, Ireland

Version of record first published: 25 Sep 2007.

To cite this article: P. N. Johnson-laird & Ruth M.J. Byrne (1995): A model point of view, Thinking & Reasoning, 1:4, 339-350

To link to this article: http://dx.doi.org/10.1080/13546789508251508

Rips, L.J. (1994). *The psychology of proofs: Deductive reasoning in human thinking.* Cambridge, MA: MIT Press.

Sperber, D., Cara, F., & Girotto, V. (in press). Relevance theory explains the selection task. *Cognition.*

A Model Point of View

P.N. Johnson-Laird, Princeton University, USA, and
Ruth M.J. Byrne, University of Dublin, Ireland

Commentary on "Pragmatic Reasoning With a Point of View" by Keith J. Holyoak and Patricia W. Cheng

Patricia Cheng and Keith Holyoak have played a central role in revealing the mysteries of the selection task (Wason, 1966). They argue that the mind contains "pragmatic reasoning schemas"; that is, principles such as:

If the precondition is not satisfied then the action must not be taken

and they have shown that when a selection task is likely to map onto the appropriate schema, subjects are likely to make a correct selection (Cheng & Holyoak, 1985). Likewise, Holyoak and Cheng (this issue) argue that the phenomena of "point of view" can be explained by combining the schemas governing permission and obligation. Readers new to the selection task may be puzzled by what it is, and so we will begin with a brief history (see also Evans, Newstead & Byrne, 1993).

The pioneer of formalism, Jean Piaget, wrote that if individuals have to verify whether x implies y, then they "will look in this case to see whether or not there is a counterexample x and non-y" (Beth & Piaget, 1966, p.181). The selection task confronts this beautiful hypothesis with an ugly fact. The task itself consists in asking subjects to choose which of four cards should be turned over in order to find out whether a conditional assertion is true or false. In a typical version of the task, the four cards have a number on one side and a letter on the other side, which the subjects know, and the cards are laid out on a table with A, B, 2, and 3 uppermost. The conditional to be tested is:

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

Intelligent adults tend to select the A card, or the A and the 2 card, and they fail to select the 3 card. In other words, given a rule of the form: if x then y, they select x, but conspicuously fail to select not-y. It is as though they have still to learn what counts as a counterexample to a conditional. However, when they are asked to generate a falsifying instance of a conditional, they usually construct

cases of x and not-y (see e.g. Oaksford & Stenning, 1992). A striking acceleration in intellectual development is accomplished merely by changing the content of the selection task. Robust effects occur with deontic conditionals concerning what is permissible (Cosmides, 1989; Griggs & Cox, 1982). Thus, as Griggs and Cox showed, the conditional:

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

tends to elicit the selection of the correct potential counterexamples (x: the card corresponding to a beer drinker, and not-y: the card corresponding to someone younger than 18). Cheng and Holyoak (1985) advanced their pragmatic theory in order to explain this sort of phenomenon. They argued that the conditional maps onto the pragmatic schema:

If the precondition is not satisfied (e.g. person is not over 18 years) then the action (e.g. drinking beer) must not be taken

and it, in turn, elicits the correct selection of cards.

Manktelow and Over (1991) carried out the selection task with the deontic conditional:

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

and the subjects' selections depended on whose point of view they were asked to take. The mother who laid down the law is concerned that her child does not cheat, and the subjects who had to take her point of view tended to select the cards:

did not tidy (not-x) went out to play (y)

The child is concerned that the mother does not renege on the deal, and the subjects who had to take the child's point of view tended to select the cards:

tidied the room (x) did not go out to play (not y).

Similar effects have been demonstrated by Politzer and Nguyen-Xuan (1992), who also showed that subjects with a neutral point of view tended to select all four cards.

The question is: what accounts for the effects of point of view? Our plan in what follows is, first, to analyse Holyoak and Cheng's answer; second, to describe the theory of mental models; and, third, to argue that it gives a more comprehensive answer.

PRAGMATIC SCHEMAS AND POINT OF VIEW

The central claim of Holyoak and Cheng's theory is that individuals have knowledge of the following sorts of principles governing permissions:

If the precondition is satisfied, then the action may be taken

and the following sorts of principles governing obligations:

If the precondition is satisfied, then the action must be taken.

These principles are known as "pragmatic reasoning schemas". Performance in the selection task, Holyoak and Cheng write (this issue, p. 291), "will be facilitated . . . when the stated rule has content that evokes a schema, and the correspondence between the stated rule and the schema rules is such that the latter map onto rules of standard logic". They then explain the effects of point of view in terms of mapping the conditional to schemas concerning the relevant individual's rights and others' duties towards this individual. Thus, from the mother's point of view, the conditional about tidying the room maps onto the permission schema given earlier for the child; whereas from the child's point of view, the conditional maps onto the obligation schema given earlier for the mother. How these mappings are made is not yet specified by the theory.

The theory of pragmatic reasoning schemas is ingenious, it has provoked interesting research, and it may be true. However, we do have three misgivings about it. First, it has a narrow purview. It has little psychological justification from outside the selection task. A major task for Holyoak and Cheng is to show how the theory applies to other sorts of reasoning and thinking. Second, the pragmatic reasoning schemas are stated using the modal auxiliaries "may" and "must". These verbs are systematically ambiguous, referring either to what is possible (or necessary) epistemically, for example:

It may rain tomorrow

or to what is possible (or necessary) deontically, for example:

You may smoke.

This ambiguity shows that the schemas are high-level rather than foundational. Holyoak and Cheng point out that rights and duties are interdefinable, but there is a more fundamental relation familiar to logicians:

possible = not necessarily not the case
necessary = not possibly not the case

This relation echoes the interdefinability of the existential and universal quantifiers:

some = not all are not
all = not some are not.

A natural interpretation of these relations is that at the heart of epistemic and deontic modalities are at least two distinct sets of possible states of affairs ("possible worlds")—those that are epistemically possible, and those that are deontically possible. If an event may happen, then it occurs in at least *some* member of the relevant set (epistemic or deontic); if an event must happen, then it occurs in *all* members of the set. Individuals know that certain actions are possible, that certain actions are permissible, and that certain actions create obligations. They can envisage such actions by building mental models of them. This knowledge and their knowledge of the meaning of "may" and "must", and other modal terms, provides the foundation for the principles that they can consciously articulate, e.g. "if you carry out certain acts, then you are morally committed to carrying out others". We take pragmatic reasoning schemas to be a systematic statement of such high-level principles, and our point is that they are *not* the deepest level of epistemic and deontic knowledge. A foundation is necessary along the lines that we will sketch out in the next section (see also Johnson-Laird, 1978).

Third, although people have knowledge that can be captured in pragmatic reasoning schemas, we have yet to be convinced that it is this knowledge in this format that is responsible for insight into the selection task. There is an alternative possibility to which we now turn.

THE THEORY OF MENTAL MODELS

The phenomena of the selection task cry out for Occam's razor. In our view, the theory of *mental models* provides such a parsimonious account (see e.g. Johnson-Laird & Byrne, 1991), and it goes beyond the data to make novel predictions. We will derive these predictions presently, but first we will outline the model theory.

The theory postulates that reasoning—deductive or inductive—is a process in which reasoners first represent the truth conditions of premises, and then use this representation together with their knowledge to construct mental models of the relevant situations. These models may take the form of visual images, but their critical feature is their structure. Thus, an exclusive disjunction:

There is a king or else there is an ace, but not both

calls for two alternative models (one for each possibility), which we represent in the following diagram:

k

a

where k denotes a model of the king and a denotes a model of the ace. The representation of explicit information is kept to a minimum so as not to overload working memory. These models are partially *implicit* because they do not make explicit that an ace does not occur in the first model and a king does not occur in the second model. Reasoners need to make a mental "footnote" that the first model exhausts the hands in which a king occurs and the second model exhausts the hands in which the ace occurs. (We have sometimes used square brackets to represent such footnotes in the diagrams, see Johnson-Laird & Byrne, 1991, but will forgo the practice here.) These footnotes, provided they are remembered, can be used to make models wholly explicit if necessary. Likewise, a conditional, such as:

If there is a king in the hand, then there is an ace

is initially represented by two models:

k a

...

Individuals grasp that both cards may be in the hand, but defer a detailed representation of the case where there is not a king in the hand. The ellipsis accordingly signifies a wholly *implicit* model, i.e. one that has no explicit content. Individuals need to make a mental footnote that the king is exhaustively represented in the explicit model, i.e. it cannot occur in the hands represented by the implicit model. Hence, the footnote can subsequently be used to infer that the antecedent does not occur in the implicit model:

k a

¬k

where ¬ represents negation. As aces are not exhaustively represented in the initial models, they may, or may not, occur in the hands represented by the implicit model:

k a

¬ k a

¬ k ¬ a

In this way, the initial models can be fleshed out so that they are completely explicit.

In general, if a conclusion holds in all the models of some factual premises, it describes a necessary conclusion; if it holds in most of the models, it describes a conclusion that is probable; and if it holds in at least one model, it describes a conclusion that is possible. Similarly, if a conclusion about an action holds in all the models of some deontic premises, it describes an obligation; if it holds in most of the models, it describes an action that is less than obligatory but more than permissible; and if it holds in at least one model, it describes an action that is permissible. These principles provide a foundation for factual, probabilistic, and deontic inferences.

People who have no training in formal logic appear to reason on the basis of mental models: they take longer and make more errors for inferences that call for multiple models than for inferences that call for only one model (see e.g. Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne, & Schaeken, 1992). They are also susceptible to "illusory" inferences predicted by the theory (see Johnson-Laird & Savary, 1995). For example, given the following premises:

> *Only one* of the following assertions is true:
> If there is a king in the hand, then there is an ace.
> If there is a queen in the hand, then there is an ace.

they infer that an ace is more likely to be in the hand than a king—a conclusion that follows from constructing models of the overall exclusive disjunction using implicit models:

```
k       a
    q   a
    ...
```

In fact, it is impossible for there to be an ace in the hand. If only one of the assertions is true, the other assertion must be false, and so either there is a king without an ace, or there is a queen without an ace. (If the premises are interpreted as bi-conditionals, the assertion is a tautology—and again one should not conclude that the ace is more probable.) There are a number of such illusions using different connectives, and they cannot be predicted by theories using valid formal rules of inference, because such rules cannot yield systematically invalid conclusions.

Holyoak and Cheng refer to the model theory as using "content-free model-theoretic procedures" (this issue, p.292) and as based on "content-independent reasoning procedures" (this issue, p.304, footnote 3). In fact, as befits a theory that was originally devised to explain the comprehension of discourse (see

Johnson-Laird, 1983), the model theory is neither content-free nor uses content-independent procedures. Content affects at least three distinct processes. First, it influences the initial interpretation of premises (see e.g. Byrne, 1989). Thus, as Johnson-Laird and Byrne (1991, p.46) point out, it will affect whether a conditional is interpreted as a one-way implication or a bi-conditional. Second, it affects the search for counterexamples to putative conclusions. If a conclusion is believable, reasoners are likely to accept it, but if it is unbelievable then, as Oakhill, Johnson-Laird, and Garnham (1989) have demonstrated, reasoners are likely to search more assiduously for counterexamples. Third, content affects the process of fleshing out implicit models (Johnson-Laird & Byrne, 1991, p.73). In fairness to Holyoak and Cheng, it is true that the model theory gives no detailed account of how knowledge is triggered during the process of comprehension. Nor, it might be added, does any other theory, including the pragmatic schemas theory, which does not explain how expressions in natural language trigger the pragmatic schemas. In short, the model theory assumes that semantic content and general knowledge play a critical role in reasoning—in interpreting premises, in fleshing out their interpretations, and in influencing the search for alternative models.

MENTAL MODELS, THE SELECTION TASK, AND POINT OF VIEW

In the selection task, subjects have a tendency not to construct counterexamples but to reason instead on the basis of their models of the conditional. A conditional, such as:

If there is an 'A', then there is a '2'

yields the initial models:

A 2
...

Subjects will select a card if it bears on the truth value of the conditional, and so they should select A alone or A and 2, depending on whether their mental footnotes correspond to a one-way or bi-conditional interpretation. These are the predominant selections for such neutral conditionals. The theory goes beyond the phenomenon to make a further prediction: there should be a correlation between the interpretation of the conditional (as a one-way or bi-conditional) and the pattern of selections. This correlation has been recently confirmed by Francesco Cara and Stefana Broadbent (personal communication).

Subjects should get the selection task right when they correctly construct counterexamples to the conditional. In an earlier analysis (Johnson-Laird & Byrne, 1991, p.80), we wrote:

An insightful performance may further depend on an explicit representation of what is not possible, i.e. the real impossibility given the rule [if A then 2] of:

A ¬ 2

This assumption appears to be correct. It is necessary to construct a model representing a counterexample to the conditional:

Impossible: A ¬ 2

A deontic conditional also has counterexamples, although they do not render the conditional false but rather are violations of the principle it expresses. The key prediction is accordingly:

Any experimental manipulation that leads reasoners to flesh out their models of the conditional, and, in particular, to construct an explicit model of an appropriate counterexample, should enhance performance in the selection task.

Such manipulations include the use of a procedure or a content likely to make counterexamples to the conditional salient, either by triggering specific memories or by eliciting a model in which such violations are represented explicitly. The model theory therefore accounts for the enhanced performance with realistic materials, including those of a deontic sort.

The key prediction goes beyond deontic effects and applies to any materials whatsoever (Johnson-Laird & Byrne, 1991, pp.80–81). It has recently been corroborated by four independent studies.

First, David Green and Rodney Larking (1995) asked subjects to construct counterexamples to conditionals that were not deontic, and showed that this manipulation enhanced performance of the selection task.

Second, Daniel Sperber, Francesco Cara, and Vittorio Girotto (in press) used a more indirect procedure with several sorts of materials that were not deontic. For example, they told their subjects that a certain machine generated cards according to the rule:

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

The machine went wrong and ceased to obey the rule, but it has been repaired, and the subject have to check that the job has been done properly. They are thus likely to represent the machine's potential error explicitly: A ¬2. The subjects in this condition and other similar ones were more successful in the selection task.

Third, Roberta Love and Claudius Kessler (1995) have independently obtained similar results. For example, they used the conditional rule:

If there are Xow then there must be a force field

where the Xow are strange crystal-like living organisms who depend for their existence on a force field. In a context that suggested the possibility of counterexamples—mutant Xows who can survive without a force field—the subjects carried out the selection task more accurately than in a control condition that did not suggest such counterexamples.

Fourth, Nira Liberman and Yechiel Klar (in press) have shown experimentally that apparent effects of a schema for "checking for cheaters" are better explained in terms of subjects' perception of appropriate counterexamples and of the relevance of looking for counterexamples.

The effects of point of view appear to arise from those conditional assertions that elicit a bi-conditional interpretation (see e.g. Fillenbaum, 1977). For example, the conditional:

If you tidy your room then you can go out to play

is usually taken to imply:

If you don't tidy your room then you cannot go out to play.

When the models for such a bi-conditional are fleshed out in a fully explicit way, they are as follows:

t	p
¬ t	¬ p

where t denotes tidying your room and p denotes going out to play. The particular cards that individuals choose in a selection task based on a bi-conditional will depend on what sort of counterexample they represent in explicit models. There are two potential counterexamples:

t	¬ p
¬ t	p

Hence, a proper test of the conditional calls for selecting all four cards (as a neutral point of view elicits from subjects, see Politzer & Nguyen-Xuan, 1992). If subjects construct an explicit model of the first counterexample, they will select the t and not-p cards; if they construct an explicit model of the second counterexample, they will select the not-t and p cards. What happens in Manktelow and Over's (1991) experiment, according to our account, is that subjects who interpret the conditional from the child's point of view construct the models:

t p
...

from which, by negation, they derive the counterexample:

t ¬ p

In contrast, subjects who interpret the conditional from the mother's point of view construct the models:

¬ t ¬ p
...

from which, by negation, they derive the counterexample:

¬ t p

In both cases, the subjects select the cards corresponding to their respective counterexamples. The model theory accordingly explains the phenomena of point of view in the selection task: when deontic rules are equivalent to bi-conditionals, instructions can lead to explicit models of one or other (or both) of the counterexamples.

The theory goes beyond the known results to make a further prediction. With a *factual* conditional that strongly suggests a bi-conditional interpretation, such as:

If the Greeks disarmed then the Turks disarmed

the same phenomena should occur as with the deontic conditionals. When counterexamples are made salient, subjects with a neutral point of view should tend to select all four cards. The Greek point of view, however, makes salient the counterexample:

Greeks disarmed ¬Turks disarmed

whereas the Turkish point of view makes salient the counterexample:

¬ Greeks disarmed Turks disarmed

Such a result would show that a deontic content is not essential for the subject's point of view to influence their performance in the selection task.

CONCLUSION

The model theory makes sense of the phenomena of the selection task: individuals will select those cards corresponding to a counterexample only if

they construct its explicit model. In the deontic domains explored by psychologists, the premises are open to a bi-conditional interpretation, and accordingly have two distinct sorts of counterexamples. The particular selections made by subjects will depend on which of these counterexamples they represent in an explicit model, and their point of view will affect which counterexamples they construct. There is no need to invoke pragmatic reasoning schemas in order to explain the phenomena. Moreover, pragmatic reasoning schemas cannot explain the results of the four studies using materials that were not deontic in which insight into the task was enhanced by making counterexamples more salient.

Our argument does not rule out the existence of pragmatic reasoning schemas, and indeed they may be used in the selection task. If theoretical entities should not be multiplied unnecessarily, however, then there is a need both for further empirical findings to bolster Holyoak and Cheng's theory and for an extension of the theory to deal with domains other than the selection task. The onus is not on pragmatic schemas alone, however. What is missing from the model theory (and the pragmatic theory) is a detailed explanation of how conditionals and other sorts of assertion elicit relevant general knowledge.

ACKNOWLEDGEMENT

This research was carried out, in part, with support from the James S. McDonnell Foundation.

REFERENCES

Beth, E.W., & Piaget, J. (1966). *Mathematical epistemology and psychology*. Dordrecht: Reidel.

Byrne, R.M.J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31*, 61–83.

Cheng, P.N., & 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.

Evans, J.St.B.T., Newstead, S.E., & Byrne, R.M.J. (1993). *Human reasoning: The psychology of deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Fillenbaum, S. (1977). Mind your p's and q's: The role of content and context in some uses of *and, or*, and *if*. In Bower, G.H. (Ed.), *Psychology of learning and motivation, Vol. 11*. New York: Academic Press.

Green, D.W., & Larking, R. (1995). The locus of facilitation in the abstract selection task. *Thinking and Reasoning, 1*(2), 183–199.

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

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

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

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

Johnson-Laird, P.N., Byrne, R.M.J., & Schaeken, W.S. (1992). Propositional reasoning by model. *Psychological Review, 99*, 418–439.

Johnson-Laird, P.N., & Savary, F. (1995). How to make the probable seem impossible. In J. Moore & J. Lehman (Eds.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Liberman, N., & Klar, Y. (in press). Hypothesis testing in Wason's selection task: Social exchange, cheating or task understanding. *Cognition*.

Love R., & Kessler, C. (1995). Focusing in Wason's selection task: content and instruction effects. *Thinking and Reasoning, 1*(2), 153–182.

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., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition, 31*, 117–140.

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

Politzer, G., & Nguyen-Xuan, A. (1992). Reasoning about conditional promises and warnings: Darwinian algorithms, mental models, and relevance judgements or pragmatic schemas. *Quarterly Journal of Experimental Psychology, 44*, 401–412.

Sperber, D., Cara, F., & Girotto, V. (in press). Relevance theory explains the selection task. *Cognition*.

Wason, P.C. (1966). Reasoning. In B.M. Foss (Ed.), *New horizons in psychology*. Harmondsworth, UK: Penguin.

Two and Three Stage Models of Deontic Reasoning

Mike Oaksford, University of Warwick, UK, and
Nick Chater, University of Oxford, UK

Commentary on "Pragmatic Reasoning With a Point of View" by Keith J. Holyoak and Patricia W. Cheng

Holyoak and Cheng (this issue; henceforth "H & C") provide a computational-level analysis (Marr, 1982) of deontic reasoning that corrects a probably too simplistic view of deontic rules (Cosmides, 1989) by introducing important ideas from jurisprudence. This analysis addresses the frequently cited criticism of pragmatic reasoning schema (PRS) theory that it does not account for the selection of the *not-p* and *q* cards in some versions of the thematic selection task. H & C suggest that people possess two PRSs, one the original permission schema from Cheng and Holyoak (1985) and an obligation schema derived from Politzer and Nguyen-Xuan (1992). They observe that these schemas are interdefinable because the antecedents and consequents of the rules that make them up involve rights and duties that are complementary. Rights and duties implicitly introduce two individuals who have different roles—e.g. employer and employee. By focusing on their rights, one of the individuals may interpret a rule as a permission whereas the other may interpret it as an obligation.