

To appear in: Manktelow, K. & Chung, M. (eds.) *Psychology of reasoning*. Psychology Press.

The natural history of hypotheses about the selection task: towards a philosophy of science for investigating human reasoning

Keith Stenning and Michiel van Lambalgen

December 1, 2002

Abstract

Wason's (1968) selection task was designed as a test of subjects' hypothesis testing behaviour, and Wason's early analysis of the difficulties subjects have with the task are closely related to Popper's philosophy of science. Popper's account of science is one of a progressive approach to a true theory through the sequential falsification of hypotheses refined in the light of previous falsification. Wason's early account of subjects' failure to turn the false-consequent card was in terms of failure to attempt to falsify.

Both Popper's account of science and Wason's account of the abstract selection task have lived full lives in the intervening years. Popper's philosophy of science is widely believed to be an at best partial account of science, which is especially weak in its lack of account of where hypotheses come from. Goodman (19??) has focused attention on the importance of the inductive history and theoretical embedding of concepts in the control of induction. Kuhn (19??) demonstrated the importance of the fact that science is a social activity and can only be understood as such. Wason's account of the selection task in terms of failure to falsify does not sit well with more recent evidence.

Meanwhile, the application of logical semantic techniques to understanding natural language has radically enriched our knowledge about the range of interpretative options subjects have for the materials used in selection task experiments. Even more importantly, the

semantic description of natural language is now understood as a descriptive rather than a normative exercise. Despite this understanding, psychologists, with few exceptions, continue to assume the appropriateness of the classical semantic analysis which Wason assumed to be the right yardstick for measuring subjects' performance.

It seems that both experimenter and subject are faced by an urgent need for a viable philosophy of discovery and justification. The purpose of this paper is to relate some of the developments in the various disciplines concerned, and to update our picture of the psychology of reasoning in the light of historical developments. The aim is to find a more viable philosophy of science which allows fruitful interaction between psychological experimentation and the many other methods of investigation which are required to understand human reasoning.

1 Introduction

Wason's selection task is as interesting for what the history of its study tells us about psychological experimenters, as for what it tells us about subjects' reasoning. In both cases the task is a rich source of insights. This chapter reprises some of what has been revealed under both headings in the thirty five years since Wason invented the task. The purpose of studying history is, of course, to avoid reliving those parts of it we would sooner have avoided the first time.

The initial publication of the selection task (Wason 1966) was contemporaneous with Chomsky's *Aspects of the theory of syntax* in linguistics, Montague's *English as a formal language* in semantics, Kuhn's *Structure of scientific revolutions* in the philosophy and sociology of science, and Goodman's *Fact, fiction and forecast*¹ in the philosophy of induction. In those intervening thirty five years, our conceptualisation of language in general and its cognitive implementations; of the semantics of conditional statements; and of how evidence is weighed in the adjudication of statements of regularities have been substantially transformed by these among many other landmark works.

But by that curious relativistic time warping which only academic disciplines can achieve, the speed and direction of progress in the psychological study of deductive reasoning, viewed from points outside has appeared incommensurable with these other progresses. Our contemporary understanding of natural language semantics, conditional semantics in particular, tells us that

¹At least the second edition.

there is a range of possible interpretations available for sentences of the form in which Wason chose to couch his rule, and a range of different expressions for each of these conditional relations. Our contemporary understanding of the relationship between evidence and generalisation in the process of developing and testing theory tells us that these processes are extremely knowledge rich and socially embedded, and that lawlike generalisations come with environments of possibly implicit background conditions which require complex bridging inferences for them to be brought to bear on data. Even nearer to home within cognitive psychology, thirty five years of psycholinguistics has taught us that interpretation and reasoning are highly interactive processes.

Nevertheless, psychological experiment on the selection task, with one or two honourable and instructive exceptions, has universally assumed that a single logical model of the conditional provides not only the experimenter with a yardstick of correct performance in the task, but also the subjects with their model of the meaning of Wason's rule; that the relation between evidence and rule is uniform and as Wason assumed it to be; that a particular rather simplistic interpretation of Popper's philosophy of science justifies the competence model by which performance is to be measured; and that the question of the origin of hypotheses has no bearing on how we should measure them against evidence.

The selection task has in fact been used as a weapon *against* the application of linguistic and semantic theory in understanding subjects' reasoning. Famously Johnson-Laird, in the 1970s used the variability of performance with what he claimed to be rules of the same form, to argue that human reasoning operated in virtue of content rather than of form. This call to arms against logical accounts was taken up by Cosmides' and her colleagues' 'evolutionary psychology'. The ramifications of this argument can be traced even in the most recent and various theories of selection task performance.

The historical reasons why psychology might clash with logic on this ground are well known. Psychology and logic had an acrimonious parting of the ways at the end of the 19C, usually traced to the twin influences of Frege's anti-psychologism and the invention of psychological experiment. This divorce was no doubt necessary for both disciplines to establish their identities and methods. That this divorce should prove such an enduring barrier to communication since is regrettable. Logic, semantics and most parts of psychology have changed out of all recognition in the meantime. Specifically, the field that has come to be known as 'formal semantics' applies the techniques of logic to the *descriptive* study of the meaning of natural language sentences and so constitutes a body of knowledge about what interpretations

of sentences like Wason's rule are likely to be entertained by subjects. It is the thesis of this chapter that the children need a dialogue.

2 Psychological interpretation of Popper's view of science

Popper (19??) argued against a positivist view of science which held that theories could be shown to be true by recourse to observations recorded in 'observation sentences' and measured for their implications against theoretical sentences. He proposed that there could be no easy division between observation and theoretical statements. Theories had to be tested against the evidence, found wanting, and revised in the light of their falsifications. If no falsification was in principle possible, no theory existed. The truth could be successively approached by continual falsification and revision, but might forever remain beyond our grasp. Theory, or the smaller unit of the hypothesis, could be tested against evidence, in the process of justification, but this process was to be carefully distinguished from the processes by which hypotheses came to be entertained—the processes of discovery. Although much could be said about 'the logic of justification' (and this is where Popper's efforts are expended) there is no analogous 'logic of discovery'.

This philosophy of science has been remarkably influential amongst psychologists studying reasoning, though it undergoes some transformation in crossing the disciplinary boundary. I understand that Wason himself resisted the inference that the selection task was inspired by Popper's philosophy of science and that Popper attended early presentations by Wason at UCL and denied the relevance of the selection task to his philosophical concerns (Manktelow personal communication). It is certainly true that there is a considerable gap between the psychological and philosophical uses of Popper's theory. Nevertheless, it is equally clear on the psychological side of the fence that the influence was strong, and is still very much alive.

Psychologists in studying reasoning have placed enormous emphasis on the testing of hypotheses and much less on where hypotheses come from. For a contemporary example, Politiek (20??) in a review which is far more theoretically informed than much of the literature, still can accurately characterise the field in stating that "I do not go into detail about how the idea has come to the subject's mind" p. 2.

Meantime, back in the philosophy of science, the relations between logic,

discovery and justification have received much scrutiny. The processes of justification are no longer seen as narrowly logical, even if some logical foundations may be required for the study of all these processes. Induction is not analogous to deduction, and while logics are systems for deduction there is no comparable ‘logic of induction’.

If the idea of a ‘logic of justification’ has taken a drubbing in the intervening years, what about the absence of a ‘logic of discovery’? Philosophers of science such as Hanson (1958) and the ‘friends of discovery’ had already pioneered the study of historical scientific discoveries. They showed that it was possible to make systematic studies of discovery though these should not be regarded as ‘logics’ anymore than should the processes of justification. The fact that highly personal psychological processes may be involved in discovery does not rule out that discovery also involves systematic processes, both psychological and sociological.

It is worth illustrating the relation between idiosyncratic and systematic aspects of discovery. Take even one of the most extreme cases of ‘irrational’ processes of invention such as Kekule’s famous discovery of the ring structure of the benzene molecule through a dream in which a snake seized its own tail. Notice what part the dream plays in the discovery of Kekule’s hypothesis.

Kekule and chemists of his time were trying to discover the structure of the benzene molecule from a starting point in which the other known organic carbon molecules were all structured as linear carbon chains. This generalisation that carbon compound molecules were composed of linear chains of carbon atoms with other groups hung off them seems to have been so well embedded in the scientific community that alternatives such as ring structures were unconsciously but never explicitly ruled out. Kekule’s dream appears to have dislodged this assumption and started him thinking about the possibility of a ring structure. Once this hypothesis was entertained, it could be seen to account for existing data and to guide further experiment to discriminate it from the linear hypothesis.

Dreaming may be an irrational process, but notice exactly what the dream does and doesn’t achieve in this process. It is only because there is a highly structured pre-existing space of hypotheses about the structure of benzene (that it is a linear chain of six carbon atoms with other groups attached) that Kekule could interpret his dream as indicating a radical new hypothesis (that benzene is a six-atom carbon ring). Anyone not party to the reigning hypothesis-space of the day could not have interpreted the dream in the way that Kekule did. That hypothesis space was the result of the complex state of chemical knowledge involving a large body of theory and data. There is

an essential level at which the phenomenon is systematic even when there are levels which are idiosyncratic. I do not wish to decry the importance of idiosyncratic psychological processes in creative thought, but they must be interpreted against the theoretical background in which they play their part.

Notice also that the idiosyncratic level is not unique to the processes of discovery. Other scientists may just as well have been influenced by their dreams in their invention of ways of *justifying* hypotheses. The relation between dream and hypothesis is the same in both cases. It would justify a hypothesis simply dreaming that the hypothesis was true, but it could be of great use dreaming how a proof or experimental demonstration could be achieved. Just as Kekule could not have interpreted his dream of discovery without the systematic background of the hypothesis space of the day, so a dream of a method of justification would require an interpretable relation to the field's methods of justification to be of any use.

In summary, Popper's (and other philosophers of science) rejection of a 'logic of discovery' needs to be reassessed in a landscape in which the logic of justification has been so thoroughly reappraised. One can agree with Popper that there is no 'logic of discovery' on the same grounds that there is no logic of justification. Neither process is of the kind that logic is now understood to capture. Both discovery and justification are highly systematic processes, both at the psychological and the sociological levels. Nevertheless, parts of both the processes of discovery and justification may also at times consist of highly idiosyncratic psychological processes such as dreaming. So understanding discovery is just as important for understanding science as is understanding justification. If this is the state of understanding in philosophy of science, Popper's backgrounding of discovery is still very much alive in the psychology of reasoning.

3 Incurious naturalists

What has been the impact on the psychology of reasoning of its particular interpretations of Popper's philosophy? Especially their impact on understanding of the selection task? In the selection task, a hypothesis arrives out of the blue, a message in a bottle on the beach so to speak. The subjects' task is interpreted as one of justification (or dismissal) of a rule. But I will argue that the subject is faced with a task of discovery which is just as central to our understanding of what happens in Wason's task. This task is the discovery of how the experimenter intends the task and rule to be understood.

This discovery process needs to give rise to hypotheses before any policy of testing can be designed, though discovery and testing may well take place iteratively. My argument is that problems with this necessary discovery are what drives the mental processes of subjects in the original descriptive (as opposed to the later deontic) versions of Wason's task.

Popper's distinction is also required to understand the behaviour of the experimenter's over the ensuing years. The experimenter's task *should* also be seen as consisting of one task of discovery and one of justification. Experimenters need to engage in descriptive explorations in order to discover what subjects are doing in the task and, thereafter, to engage in processes of justification (or dismissal) of the resulting explanatory hypotheses.

Psychologists interpretation of Popper as writing off discovery as beyond systematic study appears to have lead them to omit the descriptive and exploratory research which one might expect as a prelude to justifying hypotheses about how their subjects reason. The great bulk of the work in the psychology of deductive reasoning has consisted of attempts to find positive evidence for models of performance which arrive, just like the rule in the selection task, like messages in bottles on the beach. Looking for supporting evidence is, of course, a very un-Popperian activity, even if it may be justified at some stages in theoretical development by other philosophies of science.

What might one have expected scientists to do in response to Wason's publication of his original findings in the selection task? What are the obvious exploratory questions which one might have expected to come first, and to have constituted the process of discovery? Of course the philosophy of science has now taught us that all enquiry must be guided by at least some conceptualisation of the topic at hand, however unsystematic and we should not forget this necessary background. But we can still ask what would have been reasonable exploratory questions to have asked first? Remember, we are dealing with a field of investigators who consider themselves first and foremost to be empirical experimenters who might be expected to revel in the stage of observation and exploration.

I would argue that what is surprising is that rather few researchers spent rather little effort on the simple descriptive questions one might expect to have dominated the field initially. Question such as what makes the original task so hard? What range of explanations could distinguish the hard and the easy versions of the task? How do subjects construe this task? Do they all construe it the same way? What range of possible interpretations might subjects reasonably have of the rule? Is Wason's logical competence model a reasonable interpretation of what we know about the meaning of the rules

used?

There are of course honourable exceptions to my generalisation and we will consider some of their contributions to the debate. But the debate has not been organised by these descriptive questions. The debate has been organised by attempts to justify a set of extremely specific hypotheses about performance in the task. The field has, in Newell's (1973) memorable phrase, been playing twenty questions with nature, and it hasn't been winning.

First let us briefly review the outline history of hypotheses. Of all the researchers into selection task performance, Wason ranks as by far the most exploratory. He did explore sources of difficulty. One of his earliest exploratory papers (Wason & Shapiro 1971) in fact lists for future exploration most of the factors that have subsequently been shown to be significant, even down to the possibility of entertaining a statistical information-based model of the task. Wason did use alternative experimental paradigms such as Socratic tutoring (e.g. Wason 1969). But even Wason could not resist the allure of the hypothesis, and his early investigations are dominated by the hypothesis that subjects attempt to verify the rule rather than to falsify it—Wason's interpretation of the Popperian account of irrationality.

When several demonstrations of good reasoning with more 'concrete' materials were presented (among them some of Wason's own), attention shifted to the role of the familiarity of materials. Griggs & Cox (1982) showed that a drinking age rule produced excellent performance but also showed that Johnson-Laird, Legrenzi & Legrenzi's (1972) UK postal regulation did not facilitate performance unless the subject was familiar with the rule as a post office regulation. Manktelow & Evans (1979) then showed that familiarity *per se* was not sufficient to produce good performance in that perfectly familiar material was as hard as abstract material if the *relation* between antecedent and consequent was not clear.

Meanwhile Evans (1972) pursued the hypothesis that subjects were responding in the abstract task by merely choosing the cards which 'matched' antecedent and consequent of the rule. This hypothesis is of course nothing but a redescription of the modal response in the original abstract task until it is taken together with the 'negations paradigm' evidence that negations are ignored in the process of matching. Matching of a negated consequent to the positive instance card was deemed the logically correct choice. The matching hypothesis assumed that this was a correct choice for the wrong reason because matching was a superficial process unrelated to logical reasoning. Evans continued his pursuit of the matching hypothesis by looking at inspection times of cards and showing that subjects made fast decisions

about which cards to select and then spent large amounts of time considering these choices without further considerations of any others (exactly analogous pattern we are here arguing researchers show with regard to their favourite hypotheses.).

Cosmides (1989) built her evolutionary hypothesis on Griggs & Cox's observations of good performance with drinking age laws. She hypothesised that where performance was good it was achieved by cheating detector modules evolved in the Pleistocene era for policing social contracts. This is a rather specific hypothesis and Cosmides efforts have gone into finding evidence to support it. Some very obvious attempts to falsify it have been completely ignored. To fit drinking age laws into the category of social contracts, the laws had to be considerably bent. Cosmides & Tooby (1992) claimed that the drinking age law "expresses a social contract in which one is entitled to a benefit (beer) only if one has satisfied a requirement (being a certain age)" (p. 183). Undergraduate heaven! Even the hypothesis' internal coherence is dubious. For example, it is not clear why cheating detectors should not serve also to detect lying, which is a form of cheating more plausibly based on a 'social contract' than drinking age laws. Detecting lying *ought* to be enough to perform the descriptive selection task. All a subject with lying detectors has to do is to ask herself whether the source of the rule *could* be lying about each card. Yet even explicit instructions to take this stance do not help much with the task as Wason had already showed (Wason 19??). So the hypothesis is not even able to explain why the descriptive (abstract) selection task is so hard.

Johnson-Laird (1983???) ; Johnson-Laird & Byrne 19??) has pursued the hypothesis that selection task reasoning is done by applying the apparatus of mental models. Again this is a very specific hypothesis, even if less specific than is sometimes claimed: it is not clear that mental models, as a representational theory, is distinct from mental logic theory (the equivalence was exposed in Stenning 1992 and Stenning & Yule 1997). However, for present purposes of understanding the selection task, the most distinctive property of mental models theory (and mental logic theory) is the insistence that a single semantic interpretation of the 'if ... then' rule as a material implication is made by subjects. Even in the course of a paper arguing for the multiplicity of interpretations of "if ... then", Johnson-Laird & Byrne (in press) invoke a single semantic reading for their explanations of selection task performance.

This is a very strong hypothesis. Semantic accounts of natural language 'if ... then' are uniformly agreed that material implication is a difficult or impossible reading to induce. In fact a good deal of scorn from outside the

field has been poured on the selection task precisely because Wason assumed that material implication is the right competence model for the task, and this has been thought to make the task hopelessly artificial. We will return below to the significance of the artificiality of the task. Here the point is that while much effort has been expended on finding evidence to support the mental models hypothesis, little has been expended on attempting to falsify it.

Chater & Oaksford's (1994) application of a Bayesian optimal experiment framework as a competence model for the selection task made the valuable contribution of challenging the hegemony of the classical deductive competence model of the task assumed by Wason, and all other preceding authors. The Bayesian model assumes that subjects make certain plausible assumptions about the cards being a certain kind of sample from a larger population, and that subjects' likelihood of turning cards is in proportion to the information value of the cards as determined by these assumptions. Nevertheless, Chater & Oaksford's competence model represents another very strong hypothesis about selection task performance. Most distinctively, this model assumes the same semantic interpretation of 'if ... then' that all other models assume—namely the material conditional interpretation (see Stenning & van Lambalgen 2001). The focus on the statistical superstructure of the theory may divert attention from the necessary semantic foundations, but the relevant probabilities are probabilities of the truth of propositions, and those propositions have to be specified in a language with a semantics. The semantics Chater and Oaksford choose is the same material implication.

Finally, Sperber, Cara & Girotto (1995) argued that relevance theory explains subjects' selections. Relevance theory is a broadly based linguistic/cognitive theory about the pragmatics of natural languages. For relevance theory the selection task is a testing ground rather than the ground on which the theory was developed, and so the degree to which relevance theory permits or encourages exploratory research as opposed to focussing on a specific hypothesis has to be assessed on broader grounds which are beyond the scope of this paper. We certainly applaud Sperber and his colleagues' arguments that subjects' reasoning in the task should, on the whole, be seen as rational, and that the assessment of their rationality rests primarily on the pragmatics of the situation. However, as far as relevance theory is applied to the selection task, it shares with the other hypotheses described here the feature that it assumes the same semantic interpretation of the rule and the same underlying competence model of the abstract task, and it fails to relate problems that subjects have in the task to the specific interactions between semantic interpretations and the task.

Each of these approaches to the selection task adopted the extremely specific hypothesis that subjects' interpretation of the rule used was adequately characterised by a narrow class of material conditionals. Each of these approaches to the selection task are thoroughly Popperian in focussing on the justification of a hypothesis to the exclusion of asking where that hypothesis comes from and whether it is a reasonable or effective starting place for investigations into Wason's initial startling results. Each of these approaches is thoroughly un-Popperian in its search for verifying evidence and their failure to systematically attempt to falsify their considerable assumptions.

From our vantage point that sees Popper's estimation of the merits of falsification as overdone, one might have some sympathy with the approach of looking for supporting evidence for hypotheses, so these approaches may be better justified in some other philosophy of science. However, we will argue that this can hardly exonerate them from ignoring the problem of the subjects' interpretation of the materials. We return below to an alternative explanation for researchers' lack of exploration and this particular failure to falsify.

4 The Complete Angler

Although these have been the prominent approaches to the selection task there has been a sceptical undercurrent of more descriptive empirical work throughout its study. We have already mentioned Wason's own exploratory work looking for materials which make the task easy. Manktelow & Evans (1979) demonstration that familiarity, if it was a factor at all, operated on the relation between antecedent and consequent rather than on the contents of either. One might add Wason & Green's (1984) work on the reduced array selection task showing how restricting material on the visible sides of the cards to information relevant to the consequent of the rule made the task very much easier. Hoch & Tschirgi (1985) made concerted efforts to analyse and demonstrate sources of difficulty operating in the abstract selection task. Margolis' (1988) work which drew attention to the importance of how subjects interpretation of their responses also belongs here. Although there is evidence from early on that Wason was aware that subjects entertained biconditional interpretation of the rule, and failed to understand the intended reversibility of the cards, Gebauer & Laming (1997) were almost alone in empirically investigating the assumed semantic interpretations of the rule in the selection task. They used a repetition technique to show that a large

proportion of subjects' selections are consistent with one or other of four interpretations of the rule attained by combining conditional and biconditional readings of the connective, with constant and variable readings of the anaphors in Wason's rule. However, Gebauer & Laming still assumed that subjects make single interpretations of the rule and reason perfectly from their adopted interpretation, and all the interpretations investigated agree with the rest of the field that conditional and biconditional are to be read materially².

The major truly exploratory psychological work on subjects' frequencies of interpretation of conditionals and other connectives was Fillenbaum's (1978) but the variety of interpretations Fillenbaum exposed were not entertained in the selection task literature. For an example, Fillenbaum showed that a substantial number of subjects interpret "if ... then" conjunctively, but this interpretation was not taken seriously as a possibility in the selection task (see Stenning & van Lambalgen submitted for some exploration of this possibility).

What alternative was available to this history of attempts to test highly specialised hypotheses? Obviously, if there was no available alternative, then it is impolite at least to carp. But there was an alternative. The logic and semantics literature represents a huge body of descriptive work on the meanings of English sentences of the form 'If ... then' and the many other expressions of conditionality. The methodology of this literature is to systematise the intuitions of native speakers of the language in question. This methodology is notoriously weak when it comes to making fine distinctions about exactly when certain structures are used, but it has considerable strength in enumerating possible readings of natural language sentences, together with some information about kinds of contexts which enable the interpretations. One might argue that this literature had failed to find some readings because of its restrictions to the data of intuition, but it is rather implausible to argue that natural language sentences could *not* have readings which are described therein.

Is this reading of the situation in 1966 anachronistic? It might be argued that the semantics literature which self-consciously treats its aim as the *description* of natural language semantics postdates Montague's work. But it is nevertheless true that the preceding logical literature contains extensive discussion and cataloguing of the divergences between natural language con-

²Stenning & van Lambalgen 2001 present evidence that interpretation and reasoning interact iteratively, and that constant anaphor readings of the rule are invoked by memory demands that result from reasoning with a variable interpretation.

ditionals and those of propositional calculus. It is true that there is a crucial shift between a primarily prescriptive stance in which what is important is seen as the clarified reasoning specified in the calculus, and the relative complexities of natural language, to one that sees the natural language as another (range of) systems for reasoning. And it is missing this shift which is critical in psychologists subsequent course. But it is still true that the disparities between the material conditional and the interpretations plausibly imposed by subjects were well known in 1966. Indeed, it is doubly paradoxical that psychologists should happily accept an older epoch of logic's model as a prescription for their subjects' reasoning.

This acceptance of the prescriptions of logic and the simultaneous rejection of the descriptions of semantics cannot be entirely put down to ignorance on the part of the psychologists working on the selection task. The competence model that they entertained was derived from the logical literature, and the difficulties inherent in ascribing material conditional interpretations to natural language conditionals had been discovered and discussed at great length in that literature since the early 20C (Lewis 1912). Indeed, the paradoxes of material implication had permeated into introductory texts in logic before 1966. The lack of correspondence between material interpretations and law-like conditionals had received much discussion in the philosophy of science literature to which Popper's contributions belonged.

So a descriptive and exploratory approach to the selection task in 1966 very well could have entertained the kinds of exploratory questions I listed above. To put some flesh on the argument I will now give a very brief description of what such a program looks like before returning to the question why this was not the course taken by the field.

Somewhat belatedly, Stenning & van Lmablagen (2001; submitted) have been pursuing an exploratory program of reasearch designed to find out what is going on when subjects do the original selection task. They start from a consideration of the difference in the semantics of descriptive and deontic rules, and how the semantics of descriptive rules interact with the circumstances of the selection task to throw up difficulties for subjects. Using only simple model-theoretic differences between descriptive and deontic semantics, differences which have been well known since the classical Greeks and which are taught in most introductory logic classes, they show first that the task interpreted descriptively presents a whole range of problems which the task interpreted deontically does not. They then go on to generate initial exploratory but nevertheless strongly suggestive evidence, from Socratic tutoring of the kind that Wason used, to the effect that subjects do experience

the proposed problems in performing the task. They go on to present evidence from conventional non-interactive experiments to demonstrate that indeed these problems are also experienced in the original form of the task.

Of course this work was not the first to point out the semantic differences between descriptive and deontic conditionals. Manktelow & Over (1990) observed the logical differences between descriptive and deontic interpretation of rules. They showed how additional assumptions likely to be made by subjects about the intentions of the utterers of deontic rules gave rise to the so-called ‘perspective’ differences observed by those who had been arguing against logical accounts of the task. They argued from this analysis that the deontic and descriptive rules posed quite different tasks. But perhaps surprisingly, what they did not do was to use the semantic accounts available to them of descriptive conditionals to predict the set of problems that descriptive conditionals uniquely posed in the selection task.

The root semantic difference between descriptive interpretations and deontic interpretations of rules is simple. Because deontic rules are about the compliance of cases but the compliance of cases has no bearing on the status of the rules, the semantic relation between card and rule is independent for each card. Whether this drinker is obeying the law has no effect on the law, and does not interact with whether any other drinker is obeying the law. Whether this card complies with a descriptive rule *may* affect the truth of rule and *may* interact with other cards’ compliance or non-compliance in the effect it has. This is just the notorious asymmetry of compliant and non-compliant cases in their effects on truth value. Compliant cases can only make rules true in combination with other cases: Non-compliant cases can make rules false (at least under some circumstances). The descriptive semantics of generalisations is complicated.

This simple semantic difference gives rise to several problems for subjects reasoning in selection tasks with descriptively interpreted generalisations which do not arise with deontically interpreted rules. More adequate discussions are available in Stenning & van Lambalgen (2001) and Stenning & van Lambalgen (submitted).

In the descriptive but not in the deontic task, there are several semantic relations between rule and cases which are distinct, asymmetrical, but readily confused. Rules apply or fail to apply to cases. Cases comply with or fail to comply with rules. Some cases may make rules false and some sets of cases may make rules true. None of these relations are identical (or even exact inverses) and yet it is easy to confuse them.

In the descriptive but not in the deontic task, rules may be interpreted as brittle or robust with regard to exceptions. Most natural language generalisations are robust to at least some exceptions. If the rule is interpreted as robust, then it is not clear how to interpret the rest of the task.

In the descriptive but not in the deontic task, it is crucial what the domain of interpretation of the rule is taken to be. Although the original instructions tell the subject that the rule only applies to the four cards, it has often been pointed out that it is unclear whether they assimilate this instruction, but rather think of the cards as a sample from a larger population. This tendency is encouraged by and encourages robust interpretations of the rule. Such an interpretation also encouraged by the use of “if ... then” and the term ‘rule’. “If ... then” is more naturally used for making open-ended generalisations than contingent descriptions, and it is odd to describe a statement about four particular cards as a ‘rule’.

In the descriptive but not in the deontic task, there are anaphors (‘one side of the card’ ... ‘the other side’) which are intended to be interpreted as variable anaphors, but which are often interpreted as constant anaphors (equivalent to ‘the front’ ... ‘the back’). This reading changes the appropriate card choices.

Because the descriptive semantics (but not in the deontic semantics) involves sets of cards, subjects are faced with a problem about contingencies of card turnings. They are intended to imagine that they have to turn all the cards they need to turn before getting any feedback, but they are not told this. If they do not assume this, then their choices should be contingent on what they find on the backs of their previous choices. The instruction to not make unnecessary turns may exacerbate this problem.

In the descriptive but not in the deontic task, the subject is put in what may be a rather socially uncomfortable position of doubting the veracity of the experimenter. To find the rule false is to suggest the experimenter is a liar, but to find some drinkers breaking the law does not reflect on the experimenter.

In the original descriptive task but not in many of the deontic tasks, there is an extra cognitive load of reasoning from the categories ‘vowel’ and ‘even number’ to the instances ‘A’ and ‘7’.

We make no claim that this list of difficulties peculiar to the interaction between descriptive interpretation of the rule and the rest of the task is exhaustive. Some of these factors have been investigated before and some have not, but even where they have been investigated they have not generally been related systematically to the semantics of descriptive rules. Because

syntactic mood (indicative/subjunctive) is such an unreliable guide to descriptive and deontic interpretation, subjects have every reason to entertain the possibility of either kind of interpretation and to settle the issue on the basis of global fit to pragmatic and semantic constraints rather than through sentence characteristics. One sees this problem arising with experimenters' failing to note that their scenarios with indicative rules often demand deontic interpretations. When a quality inspector applies an indicatively stated rule to the items on a production line, she makes a deontic interpretation about what they *ought to be* like rather than an indicative one about what they are like (Sperber, Cara & Girotto (1995); Wason & Green 1984). Note that however many are items are duff, the regulation still remains in force— one hallmark of deontic semantics.

To point out these problems is, of course, not to demonstrate that they cause subjects any problems or account for differences between descriptive and deontic performance. Stenning & van Lambalgen 2001 sought *prima facie* exploratory evidence about what problems students have with the task by using Wason's socratic tutoring method. Although Wason first applied the technique to the problem he was strongly channelled in his interpretation of his results by his specific hypothesis that students' problem was their failure to seek falsification. Our technique of tutoring was similar to Wason's but our analysis of the data took a much more exploratory approach. We attempted to elicit subjects' views on some aspects of their interpretation of the semantics of the rules, but we also combed the transcripts of the video tapes for *any* evidence that spoke to their interpretations.

This approach brought to light plenty of evidence that subjects do encounter most of the problems listed above, and more. They misinterpret anaphors; are confused as to whether they will get feedback from early choices before making later choices; bow to the authority of the experimenter and assume the rule is true; treat the cards as a sample from a larger population; are unclear about what implications the possibility of exceptions has for their choices; and above all show frequent confusions about the many different semantic relations which hold between examples and rule. The cited papers contain extensive quotations of examples of confusions arising, traceable to one or more of the problems listed. The dialogues produce ample evidence that subjects' state of confusion is dynamic. They frequently show one difficulty while being clear on another at one point, only to reverse themselves at the next. We believe this vacillation is a good guide to the kind of processes that go on in subjects' reasoning. Subjects are *not* reporting earlier reasoning. They are producing reasoning in dialogue. But of course the dialogue interaction may affect their reasoning, and so we need further evidence that

these problems really are experienced in the standard task, and to estimate the size of their impacts on reasoning performance.

Some of these problems with the descriptive selection task are already known to cause difficulties. Using exemplars in the rule (instead of categories like ‘vowel’) enable an extra ??% of subject to make the intended choices. Gebauer & Laming 1997 estimate that about ??% of subjects adopt a constant anaphora reading, which is identified by incorrect choices. A sequence of task and instructional manipulations designed to remove other problems from the descriptive task is reported in Stenning & van Lambalgen (submitted). A two-rule task designed to decrease the likelihood of robust readings of the conditional enabled 20% more subjects to make the selections predicted by the logical competence model even though no ‘matching’ card is correct in this task. One is known to be false and one true, and subjects’ task is to select cards to find out which is which. Similarly, instructions that explicitly state that all selections must be made before any feedback is received enable a 14% more subjects to succeed in the standard descriptive task. Deflecting the social conflict of having to impute falsehood to the experimenter enables 9% more subjects to make correct selections. Domain manipulation?? Putting all the manipulations together???

Our experimental program of exploring difficulties with the original descriptive selection task is not yet complete, and from its nature is openended. It is hard to be sure that all difficulties of the descriptive task have been found, and hard to analyse what combinations of difficulties different subjects experience and how they interact. Nevertheless, the program has already accounted for a considerable part of the problems of the task. What is most startling about the approach is that it so easily unearths such a list of difficulties which are unique to the descriptive selection task, and how many of them had not been experimentally explored.

The empirical generalisation that arises most strongly from this approach is that reasoning in the descriptive selection task is inherently hard, and that almost all the materials which are easy turn out to invoke deontic readings. So for example, Sperber et al.’s scenarios designed to make false-consequent cards relevant (which are similar to those originally designed by Wason & Green 1984) invoke deontic readings. These inspection line scenarios in which an inspector examines cards to ensure a quality regulation is enforced is precisely a deontic reading. The findings strongly suggest, for another example, that what is crucial for Johnson-Laird, Legrenzi & Legrenzi’s (1972) postal regulation to facilitate reasoning is that subjects recognise that it has deontic force rather than that they are familiar with the rule. The rule is

stated indicatively and only those with background knowledge are likely to understand that it is intended to be interpreted as a deontic rule.

Cummins (1996) has also proposed that deontic reasoning is an especially easy kind of reasoning because we are evolutionarily equipped to perform it. Chater & Oaksford (1996) rebut Cummins' claims of innateness and modularity for deontic reasoning. But neither party to the debate draws attention to the simplicity of deontic semantics as it interacts with *the specific demands of the selection task*. Are we to suppose that descriptive selection task performance is carried out by an innate module once it is made easy by removing all the obscurities presented by the task?

The picture that emerges from this descriptive exploratory approach to what is happening as subjects reason in the descriptive selection task is one of subjects struggling to interpret how it is that the experimenter intends them to interpret a range of conflicting information. Finding an interpretation that fits all the task's constraints is not easy, and subjects are, on this view, for the most part, thinking in highly reasonable ways. Not that all solutions are equally good fits to all the constraints. Nor that Wason's particular chosen classical logical competence model is not an important solution—perhaps even arguably the best. And quite likely an educationally important one too.

On this view of the descriptive selection task, the task is an important one because it places subjects in a kind of communicational vacuum, deprived of most of the cues that they would normally exploit in understanding the semantics and pragmatics of communications. The rationale for this experiment is then just like the rationale for most laboratory experiments. By studying behaviour in abstracted environments it is possible to expose mental processes in a way that observations in natural environments cannot achieve. As such, the approach runs the dangers that all such laboratory approaches run—of producing behaviour which is artificial and unsystematically related to behaviour in more 'natural' environments. The defence of the results in this case is just as the defence always has to be—that the relations between what subjects do here and what they do naturally are sufficiently transparent that they can be analysed. The distance between the mental processes exposed in the Socratic dialogues and the mental processes shown to be going on in the standard non-interactive version of the task is shown to be much smaller than was widely believed. The problems that show up in the first are demonstrated to be at work in the second.

Unnaturalness is also a slippery notion. The selection task is highly representative of tasks that are prominent in late secondary and early tertiary education which combine the need for cooperative interpretation of constraints

(for example believing the background rule that there are letters on one side and numbers on the other) and adversarial testing of other parts of information from the same source. Seen in this light, these are exciting results. We can study the learning involved in subjects acquiring explicit control of the classical logical interpretation of the task, and the ramifications this has in their reasoning in other circumstances (see Stenning in press). The selection task becomes a window onto important processes in learning to think, processes which are central to late secondary and tertiary education. A whole vista of empirical explorations opens up.

Why did such an exploratory empirical program not take root much earlier? I think the only answer that can be given is one in terms of the original divorce of psychology and logic. Logic was dismissed by psychologists as a necessarily prescriptive study of reasoning, although, as we have seen, they simultaneously incorporated a particularly simple logical model as their account of how people actually interpret the experimental materials. And of course dismissing logic on the grounds of prescriptiveness is hopelessly simplistic. First, it is simplistic because without some standard, description of behaviour is impossible. Second, subjects themselves have prescriptive standards that they apply to their own reasoning processes. One has only to listen to the Socratic dialogues to hear subjects upbraiding themselves for what they judge to be lapses in their standards of reasoning. There is no way for psychology to escape the description of prescriptions.

Of course a consequence of rejecting semantic theory as a good descriptive starting point for an experimental program investigating human reasoning is the inevitable reinvention of it in other guises. Mental models theory is the clearest example. Despite inventing an alternative semantic theory of conditionals (Johnson-Laird & Byrne in press) which proposes thirteen? readings of conditionals, a theory which completely neglects the extensive literature it claims to replace, the account of the selection task still ascribes to subjects the least likely reading of the conditional known to science. When this is pointed out in criticism of arguments such as those for ‘deductive illusions’ (Johnson-Laird & Savary 1999) the authors argue that the materiality of implication cannot be the problem because the illusions arise with exclusive disjunction. But this misses the point that each of the connectives can be interpreted ‘materially’ or as having some non-truth functional element, and the exclusive disjunction generates the same choices of interpretation (Johnson-Laird, Legrenzi, Girotto & Legrenzi (2000).

5 If you so clever, how come you ain't rich?

One psychologist colleague presented with the argument that the semantics of descriptives as compared to the semantics of deontics is capable of predicting a whole range of difficulties in the selection task responded that this was a powerful case of hindsight. My response would be that temporal perspective is amazingly different in different disciplines. The characterisation of descriptive and deontic semantics at the granularity required for this exercise was achieved in classical Greek logic. Reasoning about how the world *is* was well understood to have different features than reasoning about how it *ought to be*, is *planned to be*, or we *would like it to be*. These insights are alive and well in logic teaching even today. So I would reject the charge of hindsight. The problem is not one of hindsight but one of disciplinary rejection of old insights through misinterpretation of their implications.

However, I would plead guilty to tardiness. I had known about the selection task since 1968 and even taught it regularly as part of courses on cognitive psychology since 1976. The inadequacies of the available accounts were fairly clear, as I believe my students would attest, and I always taught the topic as a scandal ripe for research. So why did I not make see how to base an empirical program of research on the semantics literature on natural language conditionals and subjects' reasoning in the selection task?

I now believe that my own mistake was to overestimate the distance between the student subjects and the semantic theories. I was too ready to believe that the task was an artificial task which failed to engage with subjects' knowledge of their language, and elicited purely superficial behaviour. It was obvious that the deontic versions of the task were simply logically different and much easier tasks, and besides the extravagant evolutionary claims founded on them seemed so ill-anchored as to be best left to natural selection. But I failed to see how the original descriptive task could be analysed as presenting a set of concrete problems to subjects. Even here, it was obvious that the robustness to exceptions of natural rules which arises from their non-truth functional interpretation had a major part to play in any adequate explanation. Similarly, problems of what subjects took to be the domain of interpretation was an obvious source of issues. I remember teaching about the application of the Raven's Paradox (and its involvement of assumptions about the size of sets in the domain) to the selection task.

My current reconstruction of my failure to connect is that I did not have sufficient faith in the subjects to believe that they were grappling with exactly the issues the theorists described, even if they were struggling without the

benefit of a conceptual framework or vocabulary for describing their problems. I took the gap between researcher and subject to be much greater than it is.

Another response that I have encountered from psychologist colleagues is that this account of the selection task in terms of the semantic complexities of descriptive conditionals is merely a negative spoiling of the existing positive psychological accounts which leaves us with no psychological theory of reasoning. This perception of the situation I also reject, though I think it is revealing about how the field came to be in its current condition. This response only makes sense if semantic and psychological accounts are seen as accounts competing for the same explanatory space. But semantic accounts do not compete with processing accounts, and this semantic account is used as the basis for making a series of predictions about unexplored factors affecting processing. There can be no processing accounts without a semantic foundation because without such a foundation there is no specification of what is processed with what result. This is why mental models theory has to reinvent semantic accounts of its own, and why the Bayesian theory has to have an underlying semantic account.

Far from competing, semantic research needs to be integrated into psychological research, and when it is integrated, it provides a radically different conceptualisation of psychological problems. Instead of subjects processing with a fixed interpretation, within a single system of reasoning, they are seen as, wittingly or unwittingly, making metalogical choices between alternative systems, and changing those choices interactively as they encounter new clashes with problem constraints. The content of processing is radically different as well as the outcome. With this change in conceptualisation comes a shift to an interest in individual differences in reasoning. Even subjects who make the same card selections can be seen to be making them for quite different reasons, and the different interpretations of rule and task reflected in different selections also become targets for explanation. A far richer cognitive psychology results from a richer view of the semantics (Monaghan & Stenning in press; Stenning in press). Once individual differences are acknowledged as a worthwhile focus of explanation, that immediately makes learning and the change of patterns of response an interesting topic, and re-connects the field to educational interests. It even suggests that deduction's primary importance in everyday life is its involvement whenever issues of interpretation come to the fore—interpretation of others' communications or of our own thoughts (see Stenning & Monaghan in press).

So why has the psychology of reasoning been impoverished by premature

hypothesis-testing research strategies? Again I think the explanation must be in terms of the perception that logic and psychology are mutually exclusive. This perception leads to the feeling that if psychology is to establish itself as a proper science of reasoning then it must present theories of reasoning which are not based on logical conceptualisations. These theories should preferably be counterintuitive, and they must explain the ‘fundamental human reasoning’ mechanism.

But all this program does is to reinvent the 19C logical program of rediscovering the laws of thought. Modern logic moved on from this program a long time ago. Logics are not mechanisms, whether fundamental or not. Logic now sees itself as the mathematics of information systems, whether artificial or natural. It provides an information system hypermarket from which one can choose an appropriate system for modelling whatever phenomenon takes ones’ interest. Logic is to human reasoning roughly what geometry is to human visual perception, the mathematics and conceptualisation which is required for any empirical research program, but something quite distinct from the empirical program itself.

6 A role for psychology of reasoning in cognitive science?

From a logical point of view, the psychology of reasoning is about how the systems which logic characterises are implemented in the mind. But as I write the sentence, I can hear the misinterpretations: images of the brain processing representations which are squiggly sentences punctuated by *modus ponens* . . . The statement should conjure no such images. Logics, and more plausibly fragments of logics, can be implemented in diagrams or feelings just as well as sentences, and the ‘rules’ of most direct interest to psychologists are the regularities in the theorem provers which are bits of non-logical apparatus, whose content may be probabilities just as much as sentences. Psychologically they may be habits of thought as likely as ‘rules of inference’. To try to overcome these preconceptions, consider one illustrative example, Cosmides’ cheating detectors, and compare the picture that emerges from a logical approach with that which Cosmides’ herself presents.

There is no reason why a logical approach has to feel any conflict with the idea that reasoning is at least partially implemented in emotional reactions. Cosmides might or might not be right about the involvement of our feelings

about cheating in peoples' reasoning about social contracts. I am personally inclined to think she has the makings of a plausible story, though she does not seem to have pursued the empirical program one might expect to validate it, and does not seem to have pursued the implications of the story itself very far.

It might be that at a certain stage of development/education subjects find it easy to reason with material which has the right sort of content for them to engage their feelings about cheating. It might even be that when subjects have past this stage (perhaps due to extensive education about reasoning) that what they have done is to broaden the range of materials which engage these underlying affective reactions even though these materials are now 'abstract'. Given a proposition about vowels and consonants, they might now have a strategy of asking themselves "How can I check whether the source of this sentence *could be* telling me a lie?" If they can do this, then they should be able to solve the descriptive task, at least provided they can settle all the other imponderables about the experimenter's intentions. If the statement is true then the source cannot be using it to tell a lie, though of course they might still be using it to mislead. And we know that subjects are generally well aware that a 7 with an A on the back is an exception or a counterexample to the intended reading of the rule.

A logically based psychology of reasoning would expect intimate interactions between content of materials (and interpretation of the task generally) and the logical forms (semantic interpretations) subjects assign to sentences. In fact logic is exactly a theory about how non-logical assumptions (domain specific assumptions) interact with the generalities of language to determine what is valid reasoning in particular contexts. If affective reactions play some role in subjects' reasoning in these situations, then their status in theory is as implementations of logical systems in subjects' minds.

Some psychologists' reaction to this argument is that if logic with all its formal squiggles and inference rules is only a specification of what has to be implemented, then once more we can forget about logic and just study the affective implementation. But this is to miss the point of cognitive theories of mind. Cognitive theory is always about relations between a systemic level and an implementational level. Just studying the cheating detectors has lead to the missing of the most powerful behavioural regularity in these observations—that humans have an intuitive understanding of the difference between reasoning about how the world *is* and reasoning about how the world *ought to be* quite independently of the specifics of their cheating detectors, and the selection task makes on easy and the other hard.

Further psychological investigation of the deontic vs. descriptive distinction could still benefit from the guidance of logical insight. Current developmental work on the relation between deontic reasoning and theory-of-mind abilities raises the interesting question whether it is the mental topic of reasoning about other minds that make false-belief tasks hard, or whether it is the counterfactuality that is the source of difficulty. Logically, deontic reasoning is one species of modal reasoning, though there are many species of modal reasoning which are not deontic, for example reasoning about the possibility and necessity of situations. For example, Peterson & Riggs (1999) provide some evidence that the difficulty of false-belief tasks may be more to do with modality than mentality. Again age-old logical knowledge could provide valuable guidance to empirical programs, if it were not seen as an alternative but a foundation.

7 Conclusion

Where does all this leave Popper and the impact of his teachings on our understanding of subjects and experimenters? Popper's advice had two strands: the focus on justification rather than discovery, and the invocation to attempt to falsify hypotheses rather than to seek supporting evidence. To take experimenters first, they do seem to have focussed on hypotheses with little attention to where hypotheses originate from or how they relate to other bodies of knowledge about reasoning. But this strand of Popper's philosophy has been least supported by developments in the philosophy of science. Understanding how hypotheses are embedded in existing knowledge plays a vital role in understanding where they come from, which are worth pursuing, and how they can best be investigated. The result has been a focus on over-specific hypotheses and a neglect of the exploratory and descriptive research which must play such an important part in developing sciences, and which makes up such an important contribution of empirical work.

As far as falsification of these hypotheses goes, this principle seems to have been more honoured in the breach than in compliance. The preponderance of research has sought supporting evidence for experimenter's over-specific hypotheses. We have argued here that the explanation for the acceptance of an extremely specific hypothesis about the semantics of rules, as well as all the even more specific superstructures built on that acceptance, can only be explained by a rejection of logical insight as inherently competitive with psychological explanation.

If we are to capture the thirty five year interlude in scientific research constituted by research on Wason's selection task in a philosophical framework, it seems that Kuhn's theories are much more readily applicable than Popper's. It has been the researchers' insistence on the divorce of logical and psychological research, an essentially socially based segregation of disciplines, that has been the most powerful force in shaping the course of research. Insights straddling the disciplinary boundary have been steadfastly ignored just as Kuhn's analyses of historical episodes in science show that available counterevidence is always ignored unless its relevance is socially supported.

How about the subjects—all those anonymous students who have contributed their time to the progress of science. What light has the thirty five years of research on the selection task thrown on the relevance of the philosophy of science to understanding their mental processes? My argument here has been that Wason's task above all has to be understood as a communication between experimenter and subject, though one that is cryptic in the extreme. It is the cryptic nature of the communication that invokes the effort after meaning that we observe in our Socratic dialogues, and these dialogues appear to be rather good guides to the mental processes taking place in the standard task. The intriguing observation is that experimenters' have continued for the most part unaware of the multiplicity of things they might reasonably be taken to mean by their instructions, and generations of students have been persuaded to have the insight into the intended competence model without generally appreciating the reasonableness of their initial confusions. So the competence model has been handed from generation to generation, so to speak, without much challenge as to its uniqueness or adequacy—just what Kuhn would predict during periods of 'normal science'? But what Kuhn did not pay so much attention to is the effective insulation achievable between scientific communities by the erection of discipline boundaries.

8 References

- Chater, N. & Oaksford, M. (1994) A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101:608–631.
- Chater & Oaksford (1996) Deontic reasoning, modules and innateness: a second look. *Mind and Language*, 11(2), 191–202.
- Chomsky, N. (1965) *Aspects of the theory of syntax*. MIT Press
- 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.

Cosmides and Tooby (1992) Cognitive adaptations for social exchange. In J. Barkow, Cosmides, L. and Tooby, J. (eds) *The adapted Mind: evolutionary psychology and the generation of culture*. pps. 163–228
NY: OUP

Cummins, D. (1996) Evidence for the innateness of deontic reasoning. *Mind and Language*, 11, 160–190

Evans, J. (1972) Interpretation and ‘matching bias’ in a reasoning task. *Quarterly Journal of Experimental Psychology*, 24, 193–9.

Fillenbaum, S. (1978) How to do some things with if. In Cotton and Klatzky (eds.), *Semantic functions in cognition*. Lawrence Erlbaum Associates.

Goodman, N. (1965) *Fact, fiction and forecast* 2nd ed. Indianapolis : Bobbs-Merrill.

Griggs & Cox (1982) The elusive thematic materials effect in Wason’s selection task. *British Journal of Psychology*, 73, 407–20.

Hanson, N. R. (1958) *Patterns of discovery: an inquiry into the conceptual foundations of science*. Cambridge University Press.

Hoch, S. & Tschirgi, J. (1985) Logical knowledge and cue redundancy in deductive reasoning. *Memory and Cognition*, 13 453–476.

Kuhn, Thomas (1962) *The structure of scientific revolutions*. Chicago UP.

Gebauer, G. & Laming, D (1997) Rational choices in Wason’s selection task. *Psychological Research*, 60:284–293.

Lewis, C. I. (1912) Implication and the algebra of logic. *Mind*, 21, ???–???

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

Johnson-Laird, P. & Byrne, R. (in press) *Conditionals: a theory of meaning, pragmatics and inference*. *Psychological Review*

- Johnson-Laird P. & Savary, F. (1999) *Illusory inferences: a novel class of erroneous deductions*. *Cognition* 71(3), 191-229
- Johnson-Laird, P., Legrenzi, P., Girotto, V. & Legrenzi, M. (2000) *Illusions in reasoning about consistency*. *Science*, 288 531-532.
- Manktelow, K. & Evans, J. (1979) *Facilitation of reasoning by realism: effect or non-effect?* *British Journal of Psychology*, 70, 477-88.
- Manktelow, K. & Over, D. (1990) *Inference and understanding: a philosophical perspective*. London: Routledge.
- item[] Margolis, H. (1988) *Patterns, thinking and cognition*. University of Chicago Press.
- Montague, R. (1974) *English as a formal language*. In *Formal Philosophy: Selected Papers of Richard Montague*. Thomason, R. (ed.) 108-221
- Newell, A. (1973) *You can't play twenty questions with nature and win*. In Chase, W.C. (Ed.) *Visual Information Processing*. New York. Academic Press.
- Peterson, D. M. & Riggs, K. (1999) *Adaptive modelling and mindreading*. *Mind and Language*, 14 80-112.
- Poletiek, F. (2001) *Hypothesis-testing behaviour*. Hove: Psychology Press.
- Popper, K. (1963) *Conjectures and refutations: the growth of scientific knowledge*. London : Routledge & K. Paul.
- Sperber, D., Cara, F. & Girotto, V. (1995) *Relevance theory explains the selection task*. *Cognition*, 57 31-95.
- Stenning, K. & van Lambalgen, M. (2001) *Semantics as a foundation for psychology: a case study of Wason's selection task*. *Journal of Logic, Language and Information* 10:3 273-317.
- van Lambalgen & Stenning (submitted) *A little logic goes a long way: basing experiment on semantic theory in the cognitive science of conditional reasoning*.
- Stenning, K. & Yule, P. (1997) *Image and language in human reasoning: a syllogistic illustration*. *Cognitive Psychology*, 34, pps. 109-159.

Wason, P. (1966) *Reasoning*. In Foss, B. (ed.) *New horizons in Psychology*. Harmondsworth: Penguin

Wason, P. (1969) Regression in reasoning? *British Journal of Psychology*, 60, 471–80.

Wason — ‘turn over those cards might show rule is a lie’???

Wason, P. C. & Shapiro, D. (1971) Natural and contrived experience in a reasoning paradigm. *QJEP*, 23, 63–71.

- Wason, P. & Johnson-Laird, P. (1970) A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, 1, 134–48.

Wason & Green (1984) Reasoning and mental representation. *Quarterly Journal of Experimental Psychology*, 36A, 598–611.