ELSEVIER

A little logic goes a long way: basing experiment on semantic theory in the cognitive science of conditional reasoning

Keith Stenning[*], Michiel van Lambalgen

*Human Communication Research Centre, Edinburgh University,
2, Buccleuch Place, Edinburgh EH8 9LW, UK*

Abstract

Modern logic provides accounts of both interpretation and derivation which work together to provide abstract frameworks for modelling the sensitivity of human reasoning to task, context and content. Cognitive theories have underplayed the importance of interpretative processes. We illustrate, using Wason's [Q. J. Exp. Psychol. 20 (1968) 273] selection task, how better empirical cognitive investigations and theories can be built directly on logical accounts when this imbalance is redressed.

Subjects quite reasonably experience great difficulty in assigning logical form to *descriptively* interpreted rules in this task and materials, though not to *deontically* interpreted ones. The main empirical contrasts in reasoning performance are precisely between these interpretations, so better semantic theory can explain the main empirical effects. However, a more adequate notion of logical form than that typically employed in the psychology of reasoning is required, and we offer a richer alternative.

Prima facie evidence that subjects do in fact experience the predicted difficulties in interpreting descriptive materials is provided by analyses of socratic tutoring dialogues. Experimental verification that these difficulties do in fact affect subjects' performance in the standard task is provided by six novel experimental conditions each designed to test different aspects of the semantic predictions.

The results bear out the predictions. The semantic distinction between descriptive and deontic rules interacts with the task specifics to provide powerful generalisations about reasoning which surface in detailed explanations of many disparate observations. We conclude that semantic analyses have more direct benefits for psychological investigation than is usually credited, and conversely, that the

---

[*] Corresponding author.
E-mail address: k.stenning@ed.ac.uk (K. Stenning).

extraordinary pragmatic circumstances of psychological experiments yield much thought provoking data for semantic analysis.

1. Introduction

The psychology of reasoning studies how subjects draw conclusions from premises—the process of derivation. But premises have to be interpreted before any conclusions can be drawn. Although premise interpretation has received recurrent attention (e.g., Byrne, 1989; Gebauer & Laming, 1997; Henle, 1962; Newstead, 1995) the full range of dimensions of interpretation facing the subject has not been considered. Nor has interpretation been properly distinguished from derivation *from* an interpretation in a way that enables *interactions* between interpretation and derivation to be analysed. Our general thesis is that integrating accounts of interpretation with accounts of derivation can lead to deeper insight in cognitive theory generally, and human reasoning in particular. The present paper exemplifies this general claim in the domain of Wason's (1968) selection task, an important reference point for several prominent cognitive theories of reasoning.

What is meant by interpretation in this context? Interpretation maps representation systems (linguistic, diagrammatic, . . .) onto the things in the world which are represented. Interpretation decides such matters as: which things in the world generally correspond to which words; which of these things are specifically in the domain of interpretation of the current discourse; which structural description should be assigned to an utterance; which propositions are assumed and which derived; which notions of validity of argument are intended; and so on. Natural languages such as English sometimes engender the illusion that such matters are settled by general knowledge of the language, but it is easy to see that this is not so. Each time a sentence such as "The presidents of France were bald" is uttered, its users must decide, for example, who is included, and how bald is bald, *for present purposes*. In the context of the selection task we shall see that there are quite a few such decisions which subjects have to make, each resolvable in a variety of ways, and each with implications for what response to make in the task.

Of course, interpretation, in this sense, is very widely studied in philosophy, logic and linguistics (and even psycholinguistics) as we document in our references throughout the paper. Our thesis is that interchange between these studies and psychological studies of reasoning has been inadequate. Perhaps because the methods of the fields are so divergent, there has been a reluctance to take semantic analyses seriously as a guide to psychological processes, and many of the concepts of logic are loosely employed in psychology, at best. There are, of course, honourable exceptions which we will consider in our discussion.

The current paper is part of a more general program for raising the prominence of interpretative processes in cognitive theories of human reasoning. Stenning (2002) explores the interpretative processes of representation selection that are revealed by comparing reasoning with diagrammatic versus sentential representations. Stenning and van Lambalgen (submitted) use a non-monotonic logic to model the process of credulous interpretation whereby a hearer attempts to construct the speaker's intended model, while suspending any disbeliefs. This formal semantics for the conditional provides an account of the general kind of interpretation that subjects are most likely to begin from in their attempts to interpret the task and materials in the descriptive selection task and so underpins the informal discussion here. Stenning, Yule, and Cox (1996) and Stenning and Cox (submitted) revise and extend Newstead's (1995) attempts to model the interpretational variety exhibited by subjects faced by syllogistic reasoning tasks.

In this paper we take Wason's selection task and argue that the mental processes it evokes in subjects are, quite reasonably, dominated by interpretative processes. Wason's task is probably the most intensively studied task in the psychology of reasoning literature and has been the departure point, or point of passage, for several high profile cognitive theories: mental models theory, relevance theory (Sperber, Cara, & Girotto, 1995; Sperber & Wilson 1995), 'evolutionary psychology' (Cosmides, 1989), rational analysis (Oaksford & Chater, 1994). We will argue and present empirical evidence that each of these theories misses critical contributions that logic and semantics can make to understanding the task. For various reasons the materials of the task exert contradictory pressures for conflicting interpretations, and we argue that what we observe are subjects' various, not always very successful, efforts to resolve these conflicts. The results of our experiments expose rich individual variation in reasoning and learning and so argue for novel standards of empirical analysis of the mental processes involved.

Wason's original task was presented as follows:

> Below is depicted a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is a number on one of its sides and a letter on the other.
>
> Also below there is a rule which applies only to the four cards. Your task is to decide which if any of these four cards you *must* turn in order to decide if the rule is true.
>
> Don't turn unnecessary cards. Tick the cards you want to turn.

Rule: *If there is a vowel on one side, then there is an even number on the other side*.

Cards:

A	K	4	7

The modal response (around half undergraduate subjects) is to turn A and 4. Very few subjects (around 5–10%) turn A and 7. Wason (and the great majority of researchers up to the present) assume without considering alternatives, that correct performance is to turn the A and 7 cards only. Oaksford and Chater's (1994) inductive rational choice model was the first to challenge this assumption, by rejecting deductive models entirely—more below. Wason adopted, seemingly without awareness of alternatives, this criterion of good reasoning from a classical logical interpretation of the rule. The acceptance of classical logic as a suitable normative basis for stipulating correct performance sits oddly with Wason's and other researchers' emphasis on content as opposed to logical form as a basis for modelling human reasoning. We know (and

knew in 1968) from much linguistic and philosophical study of conditionals that logically naive undergraduate subjects should not be expected to interpret Wason's rule as a classical logical conditional (material implication). Wason also knew this, if only from Wetherick (1970). Rather than rejecting logical form as a basis for analysing this reasoning, we ask why shouldn't subjects be judged on the correctness of their reasoning *according to whatever interpretations they do reasonably adopt*? This is the line of questioning this paper pursues.

In a very real sense, Wason got his own task wrong in stipulating that there was a particular 'obviously correct' answer. By the lights of the commonest interpretation of the experimental material by undergraduate subjects as a defeasible rule robust to exceptions, the 'competence' answer would be to respond that *no* combination of card choices can falsify the rule, because any possible counterexamples are indistinguishable from exceptions. And no finite combination of choices can prove the rule is true. Alternatively, subjects with other plausible interpretations of the task and rule might reasonably want to respond that several alternative sets of cards would test the rule equally well, and this again is not an available response.

There are of course many psychological reasons why we should not expect subjects to make these kinds of responses even if they were offered as possibilities. There are strong demand characteristics and authority relations in the experimental situation, and besides, subjects are not accustomed to reflecting on their language use and lack a vocabulary for talking about and distinguishing the elementary semantic concepts which are required to express these issues. Taking interpretation seriously does not mean we thereby assume reasoning is perfect, nor that we reject classical logic as one (possibly educationally important) logical model. But the unargued adoption of classical logic as a criterion of correct performance is thoroughly anti-logical. In our discussion we review some of the stances towards logic exhibited by the prominent cognitive theories that have made claims about the selection task, and appraise them from the current viewpoint.

The empirical investigation that ensued from Wason's original experiment can be seen as a search for contents of rules which make the task 'easy' or 'hard' according to the classical competence criterion. Differences in accuracy of reasoning are then explained by various classifications of content. We argue here that by far the most important determinant of ease of reasoning is whether interpretation of the rule assigns it descriptive or deontic *form*, and we explain the effect of this interpretative choice in terms of the the many problems descriptive interpretation creates in this task setting, as contrasted with the ease of reasoning in this setting with deontic interpretations.

Descriptive conditionals describe states of affairs and are therefore true or false as those states of affairs correspond to the conditionals' content. Deontic conditionals state how matters *should* be according to some (legal) law or regulation, or preference.[1] The semantic relation between sentence and case(s) for deontics is therefore quite different than for descriptives. With descriptives, *sets* of cases may make the conditional true, or make it false. With deontics, cases *individually* conform or not, but they do not affect the status of the law (or preference, or whatever). This is of course a crude specification of the distinction. We shall have some more specific proposals to make below. But it is important for the empirical investigation to focus on these blunt differences that all analyses of the distinction agree on. Our experiments do not seek to resolve fine differences between semantic analyses, but to show the empirical importance of the broad semantic categories.

In English, the semantic distinction between descriptives and deontics is not reflected simply on the surface of sentences. Deontics are often expressed using subjunctives or modals—*should, ought, must*—but are equally often expressed with indicative verbs. It is impossible to tell without consultation of context, whether a sentence such as "In the UK, vehicles drive on the left" is to be interpreted descriptively or deontically—as a generalisation or a legal prescription. Conversely, subjunctive verbs and modals are often interpreted descriptively (e.g., in the sentence "If it's 10.00 a.m., that should (must) be John", said on hearing the doorbell, the modal expresses an inference about a description). This means that we as experimenters cannot determine this semantic feature of subjects' interpretation of conditionals simply by changing auxilliary verbs in rules. A combination of rule, content, and subjects' knowledge influences whether they assign a deontic or descriptive *form*.

For example, from the selection task, when the original 'abstract' (versus, descriptive) form of the selection task proved so counter-intuitively hard, attention rapidly turned to finding materials that made the task easy. Johnson-Laird, Legrenzi, and Legrenzi (1972) showed that a version of the task using a UK postal regulation ("If a letter has a second class stamp, it is left unsealed") produced near-ceiling performance. Though they described the facilitation in terms of *familiarity*, we believe that what was critical was that the rule, though stated indicatively, was interpreted deontically by their knowledgeable subjects. The same rule was later found by Griggs and Cox (1982) to *fail* to facilitate the performance of American subjects unfamiliar with the postal regulation. Again we believe that this was because such subjects, lacking contextual knowledge, did not interpret the rule deontically but as a descriptive generalisation. What is critical, we argue, is not familiarity per se but deontic interpretation.

Wason and Green (1984) similarly showed that a rule embedded in a 'production-line inspection' scenario also produced good performance. This rule was also deontic—about what manufactured items *ought* to be like (e.g., "If the wool is blue, it must be 4 ft. long"). Griggs and Cox (1982) showed that reasoning about a drinking age law was also easy. Cheng and Holyoak (1985) developed the theory that success on deontic selection tasks was based on *pragmatic reasoning schemas*. Although they present this theory as an alternative to logic-based theories it arguably presents a fragmentary deontic logic with some added processing assumptions (theorem prover) about the 'perspective' from which the rule is viewed. However, Cheng and Holyoak did not take the further step of analysing abstractly the contrasting difficulties which descriptive conditionals pose in this task.

Cosmides and her collaborators (Cosmides, 1989; Cosmides & Tooby, 1992) went on to illustrate a range of deontic materials which produce 'good' reasoning, adding the claim this facilitation only happened with social contract rules. Cosmides and her collaborators used the argument that only social contract material was easy, to claim that humans evolved innate modular 'cheating detector algorithms' which underpin selection task performance on social contract rules. Recent work has extended the evolutionary account by proposing a range of detectors beyond cheating detectors which are intended to underpin reasoning with, for example, precautionary conditionals (Fiddick, Cosmides, & Tooby, 2000). Cummins (1996) has argued against this proliferation that the innate module concerned is more general and encompasses all of deontic reasoning. Our logical analysis of the selection task will show that once close attention is paid to the logical differences between descriptive and deontic tasks, none of this evidence can bear either way on arguments about innateness or evolution. The reasoning

task with descriptives is simply harder than that with deontics because it engenders complex conflicts of interpretation in the context of the selection task.

These observations of good reasoning with deontic conditionals and poor reasoning with descriptive conditionals were not classified as such in this literature at the time. They were rather reported as effects of *content* on reasoning with rules of the same *logical form*. Cosmides and Tooby are explicit about logic being their target:

> On this view [the view Cosmides and Tooby attack], reasoning is viewed as the operation of content-independent procedures, such as formal logic, applied impartially and uniformly to every problem, regardless of the nature of the content involved. (Cosmides & Tooby, 1992, p. 166)

Johnson-Laird equally does not allow that content can affect inference through interpretation's effect on form:

> [F]ew select the card bearing [7], even though if it had a vowel on its other side, it would falsify the rule. People are much less susceptible to this error of omission when the rules and materials have a sensible content, e.g. when they concern postal regulations . . . . Hence the content of a problem can affect reasoning, and this phenomenon is contrary to the notion of formal rules of inference. (Johnson-Laird, 1993, p. 225)

Wason himself, discussing the Wason and Shapiro Manchester/train thematic problem, rejects the idea that there are structural differences between thematic and abstract tasks:

> The thematic problem proved much easier than a standard abstract version which was *structurally equivalent* . . . . (Wason, 1987, p. 643; emphasis added)

Neither were implications of the difference in tasks ("test whether the rule is true or false" vs. "find out whether any cases are breaking the rule") ever systematically explored.

Our proposal about the selection task at its simplest is that, under descriptive interpretations, multiple asymmetrical relations between *sets* of cases play roles in determining the truth value of the rule, and it is not even clear whether the compliance or non-compliance of cases alone can make the rule, as interpreted, true or false. Under deontic interpretation, in contrast, the relation between each case and the rule is independent of the relation between other cases— cases comply or not. Case compliance has no impact on the status of the rule.

These blunt semantic differences mean that the original descriptive (abstract) task poses many difficulties to naive reasoners not posed by the deontic task. Previous work has pointed to the differences between the deontic and descriptive tasks (e.g., Oaksford & Chater, 1994; Manktelow & Over, 1990). What is novel here is the derivation of a variety of particular difficulties to be expected from the interaction of semantics and task, and the presentation of an experimental program to demonstrate that subjects really do experience these problems. Deriving a spectrum of superficially diverse problems from a single semantic distinction supports a powerful empirical generalisation about reasoning in this task that had been missed, and an explanation of why that generalisation holds. It also strongly supports the view that subjects' problems are highly variable and so reveals an important but much neglected level of empirical analysis.

It is important to distinguish coarser from finer levels of semantic analysis in understanding our predictions. At finer levels of analysis, we will display a multiplicity of interpretative choices and insist on evidence that subjects adopt a variety of them—both across subjects and within a single subject's episodes of reasoning. At this level we certainly do not predict any specific interpretation. At coarser levels, of analysis such as between truth-functional and non-truth functional conditionals, and between descriptive and deontic conditionals, it is possible to predict highly specific consequences of adopting one of the other kind of reading in different versions of the task, and to show that these consequences are evident in the data. If they do appear as predicted, then that provides strong evidence that interpretative processes are driving the data. In fact, many of these consequences have been observed before but have remained unconnected with each other, and not appreciated for what they are—the various consequences of a homogeneous semantic distinction. Semantics supplies an essential theoretical base for understanding the psychology of reasoning.

The plan of the paper is as follows. We begin by presenting in the next section what we take to be essential about a modern logical approach to such cognitive processes as are invoked by the selection task. The following section then uses this apparatus to show how the logical differences between descriptive and deontic selection tasks can be used to make predictions about problems that subjects will have in the former but not the latter. The following section turns these predictions into several experimental conditions, and presents data compared to Wason's original task as baseline. Finally, we discuss the implications of these findings for theories of the selection task and of our interpretative perspective for cognitive theories more generally.

1.1. The cognitive application of modern logic

The selection task is concerned with reasoning about the natural language conditional 'if . . . then'. The reasoning patterns that are valid for this expression can only be determined after a *logical form* is assigned to the sentence in which this expression occurs. The early interpretations of the selection task all assumed that the logical form assigned to 'if . . . then' should be the connective $\rightarrow$ with the semantics given by classical propositional logic. We want to argue that this easy identification is not in accordance with a modern conception of logic. By this, we do not just mean that modern logic has come up with other competence models beside classical logic. Rather, the easy identification downplays the complexity of the process of assigning logical form. In a nutshell, modern logic sees itself as concerned with the mathematics of reasoning systems. It is related to a concrete reasoning system such as classical propositional logic as geometry is related to light rays. It is impossible to say a priori what is the right geometry of the physical world; however, once some coordinating definitions (such as 'a straight line is to be interpreted by a light ray') have been made, it is determined which geometry describes the behaviour of these straight lines, and hypotheses about the correct geometry become falsifiable. Similarly, it does not make sense to determine a priori what is the right logic. This depends on one's notion of truth, semantic consequence, and more. But once these parameters have been fixed, logic, as the mathematics of reasoning systems, determines what is and isn't a valid consequence. In this view it is of prime importance to determine the type of parameter that goes into the definition of what a logical system is, and, of course,

the psychological purposes that might lead subjects to choose one or another setting in their reasoning. This parameter-setting generally involves as much reasoning as does the reasoning task assigned to the subject. We are thus led to the important distinction between

reasoning *from* an interpretation.

and

reasoning *for* an interpretation.

The former is what is supposed to happen in a typical inference task: given premises, determine whether a given conclusion follows. But because the premises are formulated in natural language, there is room for different logical interpretations of the given material and intended task. Determining what the appropriate logical form is in a given context itself involves logical reasoning which is far from trivial in the case of the selection task. So what are the important parameters in a logical system? Motivated by the great variety of logical systems, logicians have tried to come up with a general definition which encompasses them all. Two main approaches can be distinguished here, one syntactic, and one semantic. On the syntactic approach (for which see Gabbay, 1993), a logical system is defined by a derivability relation $\vdash$ between sentences satisfying certain minimal properties, such as, for example, that $\varphi \vdash \varphi$. This view captures a great many logical systems, based on vastly different principles. However, the *communis opinio* is that the syntactic approach is still not general enough. Take the example of the inference $\varphi \vdash \varphi$ (called *Identity* or *Reflexivity*), which is generally considered to be a minimum requirement for a logical system. Semantically speaking, it expresses that one considers the same type of models both on the left side and the right side of the turnstile. But there exist logics for which this does not apply, e.g. logics where the models appearing on the right side are the result of operations applied to models on the left side. This shows that a syntactic characterisation of logic is likely to be artificial; intuitions reside at the semantic level. The more general notion of 'logical system' is therefore semantic, in the sense that it involves the interplay between a language and its interpretation.[2] Let $\mathcal{N}$ be (a fragment of) natural language. Assigning logical form to expressions in $\mathcal{N}$ at the very least involves[3]

1. $\mathcal{L}$ a formal language into which $\mathcal{N}$ is translated
2. the expression in $\mathcal{L}$ which translates an expression in $\mathcal{N}$
3. the semantics $\mathcal{S}$ for $\mathcal{L}$
4. the definition of validity of arguments $\psi_1, \ldots, \psi_n/\varphi$, with premises ψ_i and conclusion φ.

We can see from this list that assigning logical form is a matter of setting parameters. For each item on the list, there are many possibilities for variation. Consider as a nontrivial example, the choice of a formal language. One possibility here is the ordinary recursive definition, which has clauses like 'if A, B are formulas, then so is $A \rightarrow B$', thus allowing for iteration of the conditional. However, another possibility is where formation of $A \rightarrow B$ is restricted to A, B which do not themselves contain a conditional. Or a language may contain two implication-like operators, one of which can be iterated and one of which can't. When interpreting an 'if ... then' in natural language a choice has to be made on which formal expression 'if ... then' is to be

mapped. In fact, the formally non-iterable conditional is in many ways a more appropriate model for the natural language conditional than the usual iterable construct. A possible rejoinder could be: 'Granted that a natural language conditional is hardly ever iterated (while keeping the same meaning), surely one is entitled to a bit of idealisation to smoothe the formal development?' The trouble is that this idealisation imposes a constraint on the semantics: as one can see from a formula such as $(p \to q) \to r$, a conditional in the antecedent of another conditional makes sense only if the former conditional can be false; otherwise the formula would just be equivalent to r. But many natural language conditionals, such as, for example, generic statements cannot be false; which is not to say that they are true in the classical sense. Below we will make a case for the hypothesis that the conditionals occurring in Wason's task are often interpreted as being of this non-iterable type.

Once one has chosen a formal language, one must provide a definition of satisfaction and truth. We will see that subjects do not automatically assume the classical definition of satisfaction and truth here; rather, they are groping to find a definition of truth which is appropriate to the context.

At the most abstract level, the semantics for a language is given by a recursively defined binary relation $x \rhd \varphi$, where φ is a formula. Different types of objects can be filled in for x, but the most prominent cases in logical theorising are (a) models (e.g., classical and modal logics), and (b) information states (e.g., dynamic logic). Models are descriptions of states of affairs or possible worlds, and information states describe the available evidence. The relation $\rhd$ can be read as 'makes true' or 'supports', where the latter reading is of course more appropriate if the left argument of $\rhd$ is an information state. The relation $\rhd$ may also contain an implicit numerical argument, indicating, say, degree of support.

Even when $\rhd$ is read as 'makes true', this should not be taken as implying that 'true' has a classical meaning here, satisfying *not-false = true*. That depends entirely on the nature of the recursive definition of $\rhd$, e.g. the clauses for the negation of a formula. Moreover, even if *not-false $\not\equiv$ true*, this does not imply that there exists a third truth value different from *true* and *false*, since the semantics might be non-truthfunctional altogether. But for some cases, mostly those in which we have only partial descriptions of the world, a three-valued logic may be appropriate. It is at this level that the important distinction between *descriptive* and *deontic* can be made. This distinction plays a prominent role in our analyses of the experimental data. Intuitively, one may say that a *descriptive* conditional can be true or false on a given domain; a single counterexample (as opposed to an exception) suffices to falsify the conditional. A *deontic* conditional has different logical properties: examples may or may not comply with the conditional, but by themselves examples cannot make the conditional true or false. The name 'deontic' derives from one characteristic use of conditionals of this type, as formalisations of norms; a violation of the norm does not thereby establish that the norm is false — indeed the latter expression makes no sense. But logically speaking something much more general is going on, as will be explained in Section 1.2.

Perhaps surprisingly, the definition of the *validity* of an argument is also an independent parameter. The classical definition of validity: 'an argument is valid if the conclusion is true in all models of the premises', is one possibility. We have already pointed out that this assumes that premises and conclusion are evaluated with respect to the same models. This however is not always the case.[4]

The classical notion of validity may also give way to a non-monotonic notion of validity, the general form of which is: 'an argument is valid if the conclusion is valid in all *preferred* models of the premises'. One concrete instance of this is so-called 'closed world reasoning', in which one assumes (roughly speaking) that all statements are false which are not forced to be true by the premises. This type of reasoning is *non-monotonic* in the sense that the addition of a premise α to a given set of premises Γ may destroy previous inferences from Γ alone. One example of such closed world reasoning is the often observed conversion of the conditional: " 'if A then B' implies 'if B then A' ".

The reader may think that the above variety is mainly due to logicians inventing new exotic but perfectly inapplicable systems. This is not so. As soon as logic left the confines of mathematics and turned to the formalisation of reasoning in natural language and of what is known as 'common sense reasoning' in AI, it was noticed that the parameter choices which worked well for mathematics, were unnecessarily restrictive in contexts closer to daily life; and also that there is no single setting which suffices for all such applications.

In a nutshell, therefore, the interpretative problem facing a subject in a reasoning task is to provide settings for all these parameters—this is what is involved in assigning logical form. It has been the bane of the psychology of reasoning that it operates with an oversimplified notion of logical form. Typically, in the psychology of reasoning assigning logical form is conceived of as translating a natural language sentence into a formal language whose semantics is supposed to be given, but this is really only the beginning: it fixes just one parameter. We do not claim that subjects know precisely what they are doing; that is, most likely subjects do not know in any detail what the mathematical consequences of their choices are. We do claim, however, that subjects worry about how to set the parameters, and below we offer data obtained from tutorial dialogues to corroborate this claim. This is not a descent into post-modern hermeneutics. This doomful view may be partly due to earlier psychological invocations of interpretational defences against accusations of irrationality in reasoning, perhaps the most cited being Henle (1962): 'there exist no errors of reasoning, only differences in interpretation'. It *is* possible however to make errors in reasoning: the parameter settings may be inconsistent, or a subject may draw inferences not consistent with the settings.

From the point of view of the experimenter, once all the parameters are fixed, it is mathematically determined what the extension of the consequence relation will be; and the hypotheses on specific parameter settings therefore become falsifiable. In particular, the resulting mathematical theory will classify an infinite number of reasoning patterns as either valid or invalid. In principle there is therefore ample room for testing whether a subject has set his parameters as guessed in the theory: choose a reasoning pattern no instance of which is included in the original set of reasoning tokens. In practice, there are limitations to this procedure because complex patterns may be hard to process. Be that as it may, it remains imperative to obtain independent confirmation of the parameter settings by looking at arguments very different from the original set of tokens. This is for instance our motivation for obtaining data about the meaning of negation in the context of the selection task (more on this below): while not directly relevant to the logical connectives involved in the selection task, it provided valuable insight into the notions of truth and falsity.

Psychology is in some ways harder once one acknowledges interpretational variety, but given the overwhelming evidence for that variety, responding by eliminating it from psychological

theory is truly the response of the drunk beneath the lamp post. In fact, in some counterbalancing ways, psychology gets a lot easier because there are so many independent ways of getting information about subjects' interpretations—such as tutorial dialogues. Given the existence of interpretational variety, the right response is richer empirical methods aimed at producing convergent evidence for deeper theories which are more indirectly related to the stimuli observed. What the richness of interpretation does mean is that the psychology of reasoning narrowly construed has less direct implications for the rationality of subjects' reasoning. What was right about the earlier appeals to interpretational variation is that it indeed takes a lot of evidence to confidently convict subjects of irrationality. It is necessary to go to great lengths to make a charitable interpretation of what they are trying to do and how they understand what they are supposed to do, before one can be in a position to assert that they are irrational. Even when all this is done, the irrational element can only be interpreted against a background of rational effort.

1.2. Logical forms in the selection task

We now apply the preceding considerations to the process of assigning a logical form to the conditional occurring in the standard selection task. Wason had in mind the interpretation of the conditional as a truthfunctional implication, which together with classical logic yields the material implication. Truthfunctional, because the four cards must decide the truth value of the conditional; classical logic, because the task is to determine *truth or falsity* of the conditional, implying that there is no other option. Furthermore, the task is to evaluate the rule with respect to the four cards, so if we denote the model defined by the four cards as $\mathcal{A}$, and the rule by φ, the task can be succinctly described as the question:

what further information about $\mathcal{A}$ must one obtain to be able to judge whether $\mathcal{A} \models \varphi$,

where $\models$ denotes the classical satisfaction relation?

All this is of course obvious from the experimenter's point of view, but the important question is whether this interpretation is accessible to the subject. Given the wide range of other meanings of the conditional, the subject must *infer* from the instructions, and possibly from contextual factors, what the intended meaning is. Reading very carefully, and bracketing his own most prominent meanings for the key terms involved, the subject may deduce that the conditional is to be interpreted truthfunctionally, with a classical algebra of truth values, hence with the material implication as resulting logical form. (Actually the situation is more complicated; see the next paragraph.) But this 'bracketing' is what subjects with little logical training typically find hard to do, and we now turn to their plight.

The subject first has to come up with a formal language in which to translate the rule. It is usually assumed that the selection task is about propositional logic, but in the case of 'abstract' rule one actually needs predicate logic, mainly because of the occurrence of the expression 'one side . . . other side'. One way (although not the only one) to formalise the rule in predicate logic uses the following predicates

1. $V(x, y)$ 'x is on the visible side of card y'
2. $I(x, y)$ 'x is on the invisible side of card y'

3. $O(x)$ 'x is a vowel'
4. $E(x)$ 'x is an even number'

and the rule is then translated as the following pair

$$\forall c(\exists x(V(x, c) \wedge O(x)) \rightarrow \exists y(I(y, c) \wedge E(y)))$$

$$\forall c(\exists x(I(x, c) \wedge O(x)) \rightarrow \exists y(V(y, c) \wedge E(y)))$$

This might seem pedantry, were it not for the fact that some subjects go astray at this point, replacing the second statement by a biconditional

$$\forall c(\exists x(I(x, c) \wedge O(x)) \leftrightarrow \exists y(V(y, c) \wedge E(y))),$$

or even a reversed conditional

$$\forall c(\exists x(V(x, c) \wedge E(x)) \rightarrow \exists y(I(y, c) \wedge O(y))).$$

This very interesting phenomenon will be studied further in Section 2.5.

For simplicity's sake we will focus here on the subjects' problems at the level of propositional logic. Suppose the subject has chosen some kind of propositional representation φ for the rule, in particular for the conditional. The subject must now decide how to formalise the task itself. If the subject heeds the instruction to determine whether the rule is true or false, she has to choose the formulation $\mathcal{A} \models \varphi$? that we gave above. But for some subjects this interpretation is not accessible because of the pragmatics of the task: is it really believable that the experimenter is in doubt about the truth value of the rule? Isn't it more likely that the experimenter (your professor!) wrote down a true statement—the more so since the background rule ('letters on one side, numbers on the other side') must also be taken to be true? Section 2.3 provides several examples of subjects with this type of reaction. Such subjects place the formal representation of the rule *to the left* of the 'validity' symbol, not to the right, as intended. In other words, they use it as a premise, not as a conclusion to be established or refuted. See, for example, subject 22 in Section 2.3 for an example of this.

Another class of subjects proceed analogously because they believe a conditional allows exceptions, and cannot be falsified by a single counterexample (see Section 2.1.5). These subjects' concept of conditional is more adequately captured by the following pair of statements

1. $p \wedge \neg e \rightarrow q$
2. $p' \wedge \neg q' \rightarrow e$

Here e is a proposition letter standing for 'exception', whose defining clause is 2. (In the second rule, we use p', q' rather than p, q to indicate that perhaps only some cards which satisfy p but not q qualify as bona fide exceptions.) Condition 1 then says that the rule applies only to nonexceptional cards. There are no clear falsifying conditions for conditionals allowing exceptions, so 1 and 2 are best viewed as premises. This of course changes the task, which is now seen as identifying the exceptions. There is a more general phenomenon at work here, which deserves a section of its own.

1.2.1. The big divide: descriptive and deontic conditionals

It was noticed early on that facilitation in task performance could be obtained by changing the abstract rule to a familiar rule such as

If you want to drink alcohol, you have to be over 18

though the deontic nature of the rule was not initially seen as important. This observation was one reason why formal logic was considered to be a bad guide to actual human reasoning. Logic was not able to explain how statements supposedly of the same logical form lead to vastly different performance—or so the argument went.

However, using the expanded notion of logical form given above one can see that the abstract rule and the deontic rule *are not of the same form*. One difference is in the structure of the models associated to deontic statements. We provide one especially simple definition of a deontic conditional; there are many variants, but this one will suffice for our purposes.

A model $\mathcal{A}$ is given by a set of 'worlds' or 'cases' W, together with a relation $R(v, w)$ on W intuitively meaning: 'v is an ideal counterpart to w'. That is, if $R(v, w)$, then the norms posited in w are never violated in v. With this understanding, we may define the semantics[5] of a deontic conditional $p \prec q$ by putting, for any world $w \in W$

$$w \rhd p \prec q \text{ iff for all } v \text{ such that } R(v, w) : v \rhd p \text{ implies } v \rhd q,$$

and

$$\mathcal{A} \rhd p \prec q \text{ iff for all } w \text{ in } W : w \rhd p \prec q.$$

The definition thus introduces an additional parameter R. This allows an interesting bifurcation in understanding the task. We first reformulate the selection task in terms of the semantics just given. Suppose for simplicity that the letters can only be A, K, and the numbers only 4,7. Define a model as follows. There are four worlds corresponding to the visible sides of the cards; denote these by A, K, 4,7. Then there are eight worlds corresponding to the possibilities for what is on the invisible side; denote these by $< A, 4$ and $K >$, $< A, 7 >$, $< K, 4 >$, $< K, 7 >$, $< 4, A >$, $< 4, K >$, $< 7, A >$ and $< 7, K >$. Intuitively, the initial set of the four $A, K, 4, 7$ worlds comprises the incomplete information states, which allow eight completions. This gives as domain W of the model twelve worlds in all. The 'supports' relation $\rhd$ is defined on W as follows. Let p be the proposition 'the card has a vowel', and q the proposition 'the card has an even number'.

Then we have

1. $A \rhd p$, $K \rhd \neg p$, p undecided on 4 and 7
2. $4 \rhd q$, $7 \rhd \neg q$, q undecided on A and K
3. $< A, 4 > \rhd p, q$, $< A, 7 > \rhd p, \neg q, \ldots, < 4, A > \rhd p, q, < 4, K > \rhd \neg p, q$, etc.

If the rule is understood descriptively and as applying to the four cards only, it is represented by a material implication, and hence it is interpreted relative to exhaustive and consistent sets of complete worlds, such as $\{< A, 4 >, < K, 4 >, < 4, K >, < 7, A >\}$, etc. In this case one may ask whether the rule is true or false on such a set of worlds.

If however the rule is read deontically, it is of the form $p \prec q$, and hence the model with domain the set W together with the predicate R is necessary. Define R on W by $R(A, < A, 4 >)$,

$R(7, < 7, K >), \neg R(A, < A, 7 >), \neg R(7, < 7, A >), R(K, < K, 4 >), R(K, < K, 7 >),$ $R(4, < 4, A >)$ and $R(4, < 4, K >)$. The structure $(W, R, \rhd)$ then satisfies $p \prec q$; that is, *no amount of evidence from card-turning can make the rule false*. Turning 7 to find A just means that $< 7, A >$ is not an ideal counterpart to 7.

This is actually a general phenomenon, which is not restricted to just conditionals. As we shall see, if one gives subjects the following variation on the selection task

There is a vowel on oneside of the cards *and* there is an even number on the other side[6]

they typically respond by turning the A and 4 cards, instead of just replying 'this statement is false of these four cards' (see below, Section 3.1.5). One reason for this behaviour is given by subject 22 in Section 2.4 below, who now sees the task as checking those cards which could still satisfy the conjunctive rule, namely A and 4, since K and 7 do not satisfy in any case. Such a response is only possible if one has helped oneself to a predicate such as R. Formally, one may define a deontic conjunction $p \sqcap q$ by putting, for all w in W,

$w \rhd p \sqcap q$ iff for all v such that $R(v, w) : v \rhd p \wedge q$.

In this case the worlds $< K, 4 >$ and $< K, 7 >$ are both non-ideal counterparts to the partial world K, and similarly for the partial world 7. In other words, no completion of K or 7 can be ideal, and therefore the subject has to turn only A and 4, to see whether perhaps *these* worlds are ideal[7].

What is interesting is that, viewed in this light, there is a difference in complexity between the descriptive and the deontic cases. In the latter case, one can determine the extension of R by checking the cards one at a time. There is no interference: whether the partial world A can be extended to an ideal world is independent of whether 7 can be so extended.

In the descriptive case there is a certain dependence between card choices. A subject may argue: 'If I turn A and find a 7, I know that the rule is false, so I do not have to select any other cards. The same argument holds for the 7. So how can I make a unique choice?' A particularly clear instance of this mental conflict is provided by subject 10 in Section 2.2. Rules which are interpreted descriptively thus present greater processing difficulties than rules which are interpreted deontically, and we contend that it is this processing difficulty which explains part of what have come to be called 'content effects'.

Above we have seen that subjects may be in doubt about the structure of the relevant model: whether it consists of cards, or of cards plus distinguished predicate. An orthogonal issue is, which set of cards should form the domain of the model. The experimenter intends the domain to be the set of four cards. The subjects may not grasp this; indeed there are good reasons why they shouldn't. Section 2.1.6 gives some reasons why natural language use suggests considering larger domains, of which the four cards shown are only a sample, and it presents a dialogue with a subject who has a probabilistic concept of truth that comes naturally with this interpretation of the domain.

This brings us to the notion of 'truth value on a model'. The experimenter intends the subject to operate with the classical algebra of truth values. However, especially when operating with conditionals, the subject tends to set this parameter differently. We have observed repeatedly (see Section 2.1.2) that subjects operate with a logic in which *not-false* is not the same as *true*. Theoretically, this logic could be one of a family of three-valued logics, where 'not-false'

includes the possibility 'undecided', or, more likely, it could be of the intuitionistic variety, where, very roughly speaking, $A \rightarrow B$ is true if there is a necessary connection (e.g., a proof) linking A and B. In the absence of further data it is hard to tell which logic is applied, if any, but it is worth noting that a conditional is often felt to have a lawlike character (see subject 13 in Section 2.1.2, and the discussion in Section 2.1.6); if that is so, the truth of the conditional cannot be established by pointing to the nonexistence of counterexamples.

So far we have discussed the semantic relation $x \rhd \varphi$ for the case where x is a model of some kind. When introducing this relation, we mentioned an alternative choice for x, namely an information state providing evidence relevant for φ. A quantitative information-theoretic approach to the selection task has been given by Oaksford and Chater (1994), who argue that turning the 4 card actually yields more information than turning the 7 card. We will not repeat their arguments here, nor the criticism we voiced in Stenning and van Lambalgen (2001), but we want to note in support of Oaksford and Chater (1994), that some subjects entertain both choices for x in $x \rhd \varphi$ simultaneously, and then decide that in the context of this task it is best to go the information-theoretic route. Section 2.2.1 contains several examples of subjects exhibiting this pattern.

This concludes our survey of what is involved in assigning logical form. We now turn to the demonstrations that subjects are indeed troubled by the different ways in which they can set the parameters, and that clearer task instructions can lead to fewer possibilities for the settings.

2. Designing experimental interventions

A formal analysis of the semantic and pragmatic complexities of task and rule can suggest origins of subjects' problems. We now take up the task of turning these hypotheses based on the semantics of the materials and tasks, into experimental manipulations. As a half-way house between semantics and controlled experiment, we report here excerpts from socratic tutorial dialogues to illustrate the kinds of problems subjects experience. Some of these excerpts were reported in Stenning and van Lambalgen (2001). Others are new observations from the same transcripts. Observational studies of externalised reasoning can provide prima facie evidence that these problems actually are real problems for subjects, although there is, of course, the possibility that externalising changes the task. Only controlled experiment can provide evidence that the predicted mental processes actually do take place when subjects reason in the original non-interactive task.

We present these observations of dialogues in the spirit of providing plausibility for our semantically based predictions. We assure the reader that they are representative of episodes in the dialogues—not one-offs. [8] But rather than turn these observations into a quantitative study of the dialogues which would still only bear on this externalised task, we prefer to use them to illustrate and motivate our subsequent experimental manipulations which do bear directly on the original task. We acknowledge that we cannot be certain that our interpretations of the dialogues are correct representations of mental processes—the reader will often have alternative suggestions. Nevertheless, we feel that the combination of rich naturalistic, albeit selective observations, with controlled experimental data is more powerful than either would be on its own. At the very least, the dialogues strongly suggest that there are multiple possible

confusions, and often multiple reasons for making the very same response, and so counsel against homogeneous explanations. Following the theory outlined in Section 1.1, we view these confusions as a consequence of subjects' trying to fix one of the many parameters involved in deciding upon a logical form. Here is a list of the problems faced by subjects, as witnessed by the experimental protocols. Illustrations will be provided below.

- What is truth?
- What is falsity?
- Pragmatics: the authority of the source of the rule
- Rules and exceptions
- Reasoning and planning
- Interaction between interpretation and reasoning
- Truth of the rule versus 'truth' of a case
- Cards as viewed as a sample from a larger domain
- Obtaining evidence for the rule versus evaluation of the cards
- Subjects' understanding of propositional connectives generally

Both the semantic issues facing the subject and the experimental manipulations we design to explore them may appear highly heterogeneous. Let us reiterate that what integrates the account is the way these apparently miscellaneous factors come together in explaining the dominant observation, namely contrast between reasoning with descriptively *versus* deontically interpreted rules. Everything stems from the difference in the basic semantic relation between rule and case under these two major classes of interpretation. Indeed some of the effects we explore have been noted before but have not been taken to be anything more than surface irritants because no overall framework has been available to relate them. A structured specification of the landscape of parameters which have to be set in order to determine logical interpretation is beyond the scope of this paper, but see Stenning and van Lambalgen (submitted) for a formalised account of the main features of the interpretation of the conditional which we take to be central to subjects' initial understanding of natural language conditionals in this task.

2.1. Subjects' understanding of truth and falsity

2.1.1. A two-rule task

In an earlier set of experiments (Stenning & van Lambalgen, 2001), we introduced a novel task of presenting two rules, instructing subjects that one is true and the other false, and asking them seek evidence to decide which is which. The rules were:

1. if there is a U on one side, then there is an 8 on the other side
2. if there is an I on one side, then there is an 8 on the other side

given the background rule that one side contains U or I, and the other side contains 3 or 8. In the tutorial version of this experiment, subjects were presented with real cards lying in front of them on the table. The cards shown were U, I, 8 and 3. We first asked subjects to select cards, then to imagine what could be on the other side, and lastly to turn all cards, after which subjects were given the opportunity to revise their earlier selection. In this case, both U and I carried an 8, 8 carried an I, and 3 a U.

The motivation for introducing this manipulation was two-fold. First, the Bayesian approach due to Oaksford and Chater postulates that in solving the standard Wason task, subjects always compare the rule given, to the unstated null hypothesis that antecedent and consequent are independent. We were thus interested in seeing what would happen if subjects were presented with explicit alternative non-null hypotheses.

The classical logical competence model specifies that correct performance is to turn just the 3 card. This card is alone sufficient to identify which rule is true and which false, and is the only such singly sufficient card. In this task, the 3 card therefore offers greatest information gain and so presents a useful exploration of the Bayesian approach independent of existing observations. Second, and more importantly from our perspective, explicitly telling the subject that one rule is true and one false, should background a number of issues concerned with the notion of truth, such as the possibility of the rule withstanding exceptions. The experimental manipulation turned out to be unexpectedly fruitful; while struggling through the task, subjects made comments very suggestive of where their difficulties lay.[9] Below we give excerpts from the tutorial dialogues which highlight these difficulties. Precisely because many semantic difficulties come to the surface in this novel task, it might lead to increased performance, and so it appears to be a good experiment to repeat in a standard format.

The tutorial experiment of which a part was described above, was preceded by a so called *paraphrase* task, in which subjects were asked to judge entailment relations between sentences involving propositional connectives and quantifiers. This task continues the classical work of Fillenbaum (1978) on subjects understandings of natural language connectives. For example, the subject could be given the sentence 'if a card has a vowel on one side, it has an even number on the other side', and then be asked to judge whether 'every card which has a vowel on one side, has an even number on the other side' follows from the given sentence. This example is relatively innocuous, but we will see below that these judgements can be logically startling.

2.1.2. The logic of 'true'

On a classical understanding of the two-rule task, the competence answer is to turn the 3; this would show which one of the rules is false, hence classically also which one is true. This classical understanding should be enforced by explicitly instructing the subjects that one rule is true and the other one false. Interestingly, some subjects refuse to be moved by this instruction, insisting that 'not-false' is not the same as 'true'. These subjects are thus guided by some nonclassical logic. Some subjects, when reading the rule(s) aloud actually inserted a modality in the conditional:
Subject 13. [Standard Wason task]

S. . . . if there is an A, then there is a 4, necessarily the 4 . . . [somewhat later] . . . if there is an A on one side, necessarily a 4 on the other side . . . .

If truth involves necessity, then the absence of counterexamples is not sufficient for truth.
Subject 17.

S. [Writes miniature truth tables under the cards.]

E. OK so if you found an I under the 3, you put a question mark for rule 1, and rule 2 is false; if you turned the 3 and found a U, then rule 1 is false and rule 2 is a question mark. So you want to turn 3 or not?

S. No.

E. Let's actually try doing it. Turn over the U, you find a 3, which rule is true and which rule is false?

S. (Long pause)

E. Are we none the wiser?

S. No, there's a question mark.

E. It could have helped us, but it didn't help us?

S. Yes.

⋮

E. OK and the 3

S. Well if there is a U then that one is disproved [pointing to the first rule] and if there is an I then that one is disproved [pointing to the second rule]. But neither rule can be proved by 3.

⋮

E. Turn over the last card [3] and see what's on the back of it . . . so it's a U. What does that tell us about the rule?

S. That rule one is false and it doesn't tell us anything about rule 2?

E. Can't you tell anything about rule 2?

S. No.

The subject thinks falsifying rule 1 does not suffice and now looks for additional evidence to support rule 2. In the end she chooses the 8 card for this purpose, which is of course not the competence answer even when 'not-false' is not equated with 'true' (the I card would have to be chosen). Here are two more examples of the same phenomenon.

Subject 8.

S. I wouldn't look at this one [3] because it wouldn't give me appropriate information about the rules; it would only tell me if those rules are wrong, and I am being asked which of those rules is the correct one. Does that make sense?

Subject 5.

E. What about if there was a 3?

S. A 3 on the other side of that one [U]. Then this [rule 1] isn't true.

E. It doesn't say . . . ?

S. It doesn't say anything about this one [rule 2].

E. And the I?

S. If there is a 3, then this one [rule 2] isn't true, and it doesn't say anything about that one [rule 1].

The same problem is of course present in the standard Wason task as well, albeit in a less explicit form. If the cards are A, K, 4 and 7, then turning A and 7 suffices to verify that the

rule is not false; but the subject may wonder whether it is therefore true. For instance, if the concept of truth of a conditional involves attributing a lawlike character to the conditional, then the absence of counterexamples does not suffice to establish truth. Let us note here that it is precisely this difficulty which is absent in the case of deontic rules such as

If you want to drink alcohol in this bar, you have to be over 18.

Such a rule cannot be shown to hold by examining cases; at most we can establish that it is not violated. So in the deontic case, subjects only have to do what they find easy in any case.

2.1.3. The logic of 'false'

Interesting things happen when one asks subjects to meditate on what it could mean for a conditional to be false. As indicated above, the logic of 'true' need not determine the logic of 'false' completely. The paraphrase task alluded to above showed that a conditional $(p \rightarrow q)$ being false, i.e. is often (>50%) interpreted as $p \rightarrow \text{not } q$! (We will refer to this property as *strong falsity*.) This observation is not ours alone: Fillenbaum observed that in 60% of the cases the negation of a causal temporal conditional $p \rightarrow q$ ('if he goes to Amsterdam, he will get stoned') is taken to be $p \rightarrow \text{not } q$; for contingent universals (such as the rule in the selection task) the proportion is 30%. In our experiment the latter proportion is even higher. Here is an example of a subject using strong falsity when asked to imagine what could be on the other side of a card.

Example Subject 26 [Standard Wason task; subject has chosen strong falsity in paraphrase task]

E. So you're saying that if the statement is true, then the number [on the back of A] will be 4. . . . What would happen if the statement were false?
S. Then it would be a number other than 4.

Note that strong falsity encapsulates a concept of necessary connection between antecedent and consequent in the sense that even counterexamples are no mere accidents, but are governed by a rule. If a subject believes that true and false in this situation are exhaustive, this could reflect a conviction that the cards have been laid out according to *some* rule. It is interesting to see what this interpretation means for card choices in the selection tasks. If a subject has strong negation but still believes true and false are exhaustive, then (in the standard Wason task) *either* of the cards p, q can show that $p \rightarrow q$ is not-false, hence true. Unfortunately, in the standard set up 'either of A, 4' is not a possible response offered. In the tutorial experiment involving the two-rule task subjects were at liberty to make such choices. In this case strong falsity has the effect of turning each of the two rules into a biconditional, 'U if and only if 8' and 'I if and only if 8', respectively. *Any* card now distinguishes between the two rules, and we do indeed find subjects emphatically making this choice:

E. OK so you want to revise your choice or do you want to stick with the 8?
S. No no . . . I might turn all of them.
E. You want to turn all of them?
S. No no no just one of them, any of them.

Perhaps the customary choice of p, q in the standard task is the projection of 'either of p, q' onto the given possibilities. Another option is that some subjects have a biconditional

reading of 'if . . . then' together with strong falsity; in this case both *p* and *q* are necessary. These considerations just serve to highlight the possibility that a given choice of cards is made for very different reasons by different subjects, so that by itself statistical information on the different card choices in the standard task must be interpreted with care.

2.1.4. Truth of the rule and 'truth of the card'

Subjects are persistently confused about several notions of truth that could possibly be involved. The intended interpretation is that the domain of discourse consists of the four cards shown, and that the truth value of the rule is to be determined with respect to that domain. This interpretation is however remarkably difficult to get at. An alternative interpretation is that the domain is some indefinitely large population of cards, of which the four cards shown are just a sample; this is the intuition that lies behind Oaksford and Chater's Bayesian approach. We will return to this interpretation in Section 2.1.6 below. The other extreme is that each card defines a domain of its own, i.e. each card is to be evaluated against the rule independently. The latter interpretation is the one suited to deontic conditionals, but there are indications that subjects sometimes impose this interpretation also in the indicative case, and then struggle with the resulting clash between two notions of truth. If a card complies with the rule, in other words 'if the rule is true of the card', then some subjects seem to have a tendency to transfer this notion of truth to 'truth of the rule *tout court*'. Here is an example of the phenomenon, observed in the two-rule task.

Subject 10.

- E. If you found an 8 on this card [I], what would it say?
- S. It would say that rule two is true, and if the two cannot be true then rule one is wrong . . . (Subject turns 8).
- E. OK so it's got an I on the back, what does that mean?
- S. It means that rule two is true.
- E. Are you sure?
- S. I'm just thinking whether they are exclusive, yes because if there is an I then there is an 8. Yes, yes, it must be that.

One experimental manipulation in the tutorial dialogue for the two-rule task addressed this problem by making subjects first turn U and I, to find 8 on the back of both. This caused great confusion, because the subjects' logic (transferring 'truth of the card' to 'truth of the rule') led them to conclude that therefore both rules must be true, contradicting the instruction.

Subject 18 [Initial choice was 8.]

- E. Start with the U, turn that over.
- S. U goes with 8.
- E. OK now turn the I over.
- S. Oh God, I shouldn't have taken that card, the first . . . .
- E. You turned it over and there was an 8.
- S. There was an 8 on the other side, U and 8. If there is an I there is an 8, so they are both true. [Makes a gesture that the whole thing should be dismissed.]

Subject 28.

E. OK turn them.
S. [turns U, finds 8] So rule one is true.
E. OK for completeness' sake let's turn the other cards as well.
S. OK so in this instance if I had turned that one [I] first then rule two would be true and rule one would be disproven. Either of these is different. [U or I]
E. What does that actually mean, because we said that only one of the rules could be true. Exactly one is true.
S. These cards are not consistent with these statements here.

On the other hand subjects who ultimately got the two-rule task right also appeared to have an insight into the intended relation between rule and cards.
Subject 6.

E. So say there were a U on the back of the 8, then what would this tell you?
S. I'm not sure where the 8 comes in because I don't know if that would make the U-one right, because it is the opposite way around. If I turned that one [pointing to the U] just to see if there was an 8, if there was an 8 it doesn't mean that rule two is not true.

We claim that part of the difficulty of the standard task involving a descriptive rule is the possibility of confusing the two relations between rule and cards. Transferring the 'truth of the card' to the 'truth of the rule' may be related to what Wason called 'verification bias', but it seems to cut deeper. One way to transfer the perplexity unveiled in the above excerpts to the standard task is to do a tutorial experiment where the A has a 4 on the back, and 7 an A. If a subject suffering from a confusion about the relation between cards and rule turns the A and finds 4, he should conclude that the rule is true, only to be rudely disabused upon turning 7. Unfortunately we haven't yet done this manipulation. In any case it is clear that for a deontic rule no such confusion can arise, because the truth value of the rule is not an issue.

2.1.5. Exceptions and brittleness

The concept of truth Wason intended is that of 'true without exceptions', what we call a brittle interpretation of the conditional. It goes without saying that this is not how a conditional is generally interpreted in real life. And we do find subjects who struggle with the required transition from a notion of truth which withstands some exceptions, to exceptionless truth.
Subject 18.

E. What could you say is on the back of the 3, are you sticking with the consonant?
S. Consonant or U.
E. OK.
S. [Turns 3 and finds U] OK . . . well no . . . well that could be an exception you see.
E. The U?
S. The U could be an exception to the other rule.

E. To the first rule?

S. Yes, it could be an exception.

E. So could you say anything about the rule based on this? Say, on just having turned the U and found a 3?

S. Well yes, it could be a little exception, but it does disprove the rule so you'd have to . . .

E. You'd have to look at the other ones?

S. Yes.

Similarly in the standard Wason task:
Subject 18.

S. If I just looked at that one on its own [7/A] I would say that it didn't fit the rule, and that I'd have to turn that one [A] over, and if that was different [i.e., if there wasn't an even number] then I would say the rule didn't hold.

E. So say you looked at the 7 and you turned it over and you found an A, then?

S. I would have to turn the other cards over . . . well it could be just an exception to the rule so I would have to turn over the A.

Clearly, if a counterexample is not sufficient evidence that the rule is false, then it is dubious whether card-turnings can prove the rule to be true or false at all. Subjects may accordingly be confused about how to interpret the instructions of the experiment. In any case a $\neg q$ card would lose some salience (if it had any to begin with).

2.1.6. The cards as sample

Above we noted that there are problems concerning the domain of interpretation of the conditional rule. The intended interpretation is that the rule applies to the four cards shown only. However, the semantics of conditionals is such that they tend to apply to an open-ended domain of cases. This can best be seen in contrasting universal quantification with the natural language conditional. Universal quantification is equally naturally used in framing contingent contextually determined statements as open-ended generalisations. So, to develop Goodman's (1954) example, "All the coins in my pocket this morning are copper" is a natural way to phrase a local generalisation with a fixed enumerable domain of interpretation. However, "If a coin is in my pocket this morning, it's copper" is a distinctly unnatural way of phrasing the same claim. The latter even invites the fantastical interpretation that if a silver coin were put in my pocket this morning it would become copper—that is an interpretation in which a larger open-ended domain of objects is in play.

Similarly in the case of the four card task, the clause that "the rule applies only to the four cards" has to be explicitly included. One may question whether subjects take this clause on board, since this interpretation is an unnatural one for the conditional. It is further unnatural to call the sentence a *rule* if its application is so local. A much more natural interpretation is that the four cards are a sample. Indeed this is the point of purchase of Oaksford and Chater's proposals that performance is driven by subjects' assumptions about the larger domain of interpretation. We do find subjects who think that truth or falsity can only be established by (crude) probabilistic considerations:

Subject 26.

S. [has turned U,I, found an 8 on the back of both] I can't tell which one is true.
E. OK let's continue turning.
S. [turns 3] OK that would verify rule two. [. . .] Well, there are two cards that verify rule two, and only one card so far that verifies rule one. Because if this [3] were verifying rule one, it should be an I on the other side.
E. Let's turn [the 8].
S. OK so that says that rule two is true as well, three of the cards verify rule two and only one verifies rule one.
E. So you decide by majority.
S. Yes, the majority suggests rule two.

It is interesting that 3/U is described as *verifying* rule two, rather than *falsifying* rule one; U→8 is never ruled out:

S. It's not completely false, because there is one card that verifies rule one.

Summarising: natural language descriptive conditionals bear complex relations to cases and sets of cases in their domain. In principle, only *sets* of cases can make a descriptive rule true. Even then the fact that all cases comply may intuitively not be enough, for instance when a subject hesitates to conclude 'true' from 'not false'. The situation is still more complex because descriptive rules usually tolerate some exceptions. To get Wason's desired interpretation of the rule as a material conditional, it is necessary to background the complex range of possibilities for descriptive rules' relations to compliant cases and to exceptions, and to induce the intended meaning of 'true' and 'false'. Here the two-rule task may have a role to play. If subjects were assured that one of two rules was false and one was true, and instructed that their task was to gather minimal evidence as to which rule was which, then this hopefully focusses their attention on the more straightforward relations between rules and cases, and backgrounds the higher-order issues about how exceptions affect the truth of rules, and more generally the nature of truth. Of course the excerpts given above have mainly illustrated subjects' difficulties in the two-rule task. However, several tutorial dialogues involving the two-rule task also showed (very gradual and faltering) progress toward insight, while this progress was absent in the dialogues involving the standard task. This gave us some confidence that the two-rule task might be helpful in reaching the competence response, a prediction borne out by the experimental results reported below.

2.2. Dependencies between card-choices

The tutorial dialogues suggest that part of the difficulty of the selection task consists in having to choose a card *without being able to inspect what is on the other side of the card*. This difficulty can only be made visible in the dialogues because there the subject is confronted with real cards, which she is not allowed to turn at first. It then becomes apparent that some subjects would prefer to solve the problem by 'reactive planning', i.e. by first choosing a card, turning it and deciding what to do based on what is on the other side. This source of difficulty is obscured by the standard format of the experiment. The form invites the subjects to think of the cards depicted as real cards, but at the same time the answer should be given on the

basis of the representation of the cards on the form, i.e. with inherently unknowable backs. The instruction 'Tick the cards you want to turn . . . ' clearly does not allow the subject to return a reactive plan. This is a pity, because the tutorials amply show that dependencies are a source of difficulty. Here is an excerpt from a tutorial dialogue in the two-rule condition.

Subject 1.

- E. Same for the I, what if there is an 8 on the back?
- S. If there is an 8 on the back, then it means that rule two is right and rule one is wrong.
- E. So do we turn over the I or not?
- S. Yes. Unless I've turned the U already.

And in a standard Wason task:

Subject 10.

- S. OK so if there is a vowel on this side then there is an even number, so I can turn A to find out whether there is an even number on the other side or I can turn the 4 to see if there is a vowel on the other side.
- E. So would you turn over the other cards? Do you need to turn over the other cards?
- S. I think it just depends on what you find on the other side of the card. No I wouldn't turn them.

⋮

- E. If you found a K on the back of the 4?
- S. Then it would be false.

⋮

- S. But if that doesn't disclude [sic] then I have to turn another one.
- E. So you are inclined to turn this over [the A] because you wanted to check?
- S. Yes, to see if there is an even number.
- E. And you want to turn this over [the 4]?
- S. Yes, to check if there is a vowel, but if I found an odd number [on the back of the A], then I don't need to turn this [the 4].
- E. So you don't want to turn . . .
- S. Well, I'm confused again because I don't know what's on the back, I don't know if this one . . .
- E. We're only working hypothetically now.
- S. Oh well, then only one of course, because if the rule applies to the whole thing then one would test it.

⋮

- E. What about the 7?
- S. Yes the 7 could have a vowel, then that would prove the whole thing wrong. So that's what I mean, do you turn one at a time or do you . . . ?

⋮

- E. Well if you needed to know beforehand, without having turned these over, so you think to yourself I need to check whether the rule holds, so what cards do I need to turn over? You said you would turn over the A and the 4.

S. Yes, but if these are right, say if this [the A] has an even number and this has a vowel [the 4], then I might be wrong in saying "Oh it's fine", so this could have an odd number [the K] and this a vowel [the 7] so in that case I need to turn them all.

E. You'd turn all of them over? Just to be sure?

S. Yes.

Once one has understood Wason's intention in specifying the task, it is easy to assume that it is obvious that the experimenter intends subjects to decide what cards to turn *before* any information is gained from any turnings. Alternatively, and equivalently, the instructions can be interpreted to be to assume the minimal possible information gain from turnings. However, the obviousness of these interpretations is possibly greater in hindsight, and so we set out to test whether they are a source of difficulty in the task. Note that no contingencies of choice can arise if the relation between rule and cards is interpreted deontically. Whether one case obeys the law is unconnected to whether any other case does. Hence the planning problem indicated above cannot arise for a deontic rule, which might be one explanation for the good performance in that case.

In this connection it may be of interest to consider the so-called *reduced array selection task*, or RAST for short, due to Wason and Green and discussed extensively by Margolis (1988). In its barest outline[10] the idea of the RAST is to remove the p and $\neg p$ cards from the array of cards shown to the subject, thus leaving only q and $\neg q$. The p and $\neg p$ cards cause no trouble in the standard task in the sense that p is chosen almost always, and $\neg p$ almost never, so one would expect that their deletion would cause little change in the response frequencies for the remaining cards. Surprisingly however, the frequency of the $\neg q$ response increases dramatically. From our point of view, this result is perhaps less surprising, because without the possibility to choose p, dependencies between card choices can no longer arise. This is not to say that this is the only difficulty the RAST removes.

2.2.1. Getting evidence for the rule versus evaluation of the cards

A related planning problem, which can however occur only on a non-standard logical understanding of the problem, is the following. In a few early tutorial dialogues involving the two-rule experiment, the background rule incorrectly failed to specify that the cards have one side either U or I and on the other side either 3 or 8, owing to an error in the instructions. In this case the competence response is not to turn 3 only, but to turn U, I and 3. But several subjects did not want to choose the 3 for the following reason.

Subject 7.

S. Then I was wondering whether to choose the numbers. Well, I don't think so because there might be other letters [than U,I] on the other side. There could be totally different letters.

E. You can't be sure?

S. I can't be sure. I can only be sure if there is a U or an I on the other side. So this is not very efficient and this [3] does not give me any information. But I could turn the U or the I.

Apparently the subject thinks that he can choose between various sets of cards, each sufficient, and the choice should be as parsimonious as possible in the sense that every outcome of

a turning must be relevant. To show that this is not an isolated phenomenon, here is a subject engaged in a standard Wason task:

Subject 5.

E. So you would pick the A and you would pick the 4. And lastly the 7?

S. That's irrelevant.

E. So why do you think it's irrelevant?

S. Let me see again. Oh wait so that could be an A or a K again [writing the options for the back of 7 down], so if the 7 would have an A then that would prove me wrong. But if it would have a K then that wouldn't tell me anything.

E. So?

S. So these two [pointing to A and 4] give me more information, I think.

E. [. . .] You can turn over those two [A and 4].

S. [turns over the A]

E. So what does that say?

S. That it's wrong.

E. And that one [4]?

S. That it's wrong.

E. Now turn over those two [K and 7].

S. [Turning over the K] It's a K and 4. Doesn't say anything about this [pointing to the rule]. [After turning over the 7] Aha.

E. So that says the rule is . . . ?

S. That the rule is wrong. But I still wouldn't turn this over, still because I wouldn't know if it would give an A, it could give me an a K and that wouldn't tell me anything.

E. But even though it could potentially give you an A on the back of it like this one has.

S. Yes, but that's just luck. I would have more chance with these two [referring to the A and the 4].

These subjects have no difficulty evaluating the meaning of the possible outcomes of turning 3 (in the two-rule task), or 7 (in the standard Wason task), but their choice is also informed by other considerations, in particular a perceived trade-off between the 'information value' of a card and the penalty incurred by choosing it. Of course this does not yet explain the evaluation of the 4/K card as showing that the rule is wrong, and simultaneously taking the K/4 card to be irrelevant. The combined evaluations seem to rule out a straightforward biconditional interpretation of the conditional, and also the explanation of the choice of 4 as motivated by a search for confirmatory evidence for the rule, as Wason would have it. This pattern of evaluations is not an isolated phenomenon, so an explanation would be most welcome. Even without such an explanation it is clear that the problem indicated, how to maximise information gain from turnings, cannot play a role in the case of deontic conditionals, since the status of the rule is not an issue.

2.3. The pragmatics of the descriptive selection task

The descriptive task demands that subjects seek evidence for the truth of a statement which comes from the experimenter. The experimenter can safely be assumed to know what is (or

is deemed to be) on the back of the cards. If the rule is false its appearance on the task sheet amounts to the utterance, by the experimenter, of a knowing falsehood, possibly with intention to deceive. It is an active possibility that doubting the experimenter's veracity is a socially uncomfortable thing to do.

Quite apart from possible social psychological effects of discomfort, the communication situation in this task is bizarre. The subject is first given one rule to the effect that the cards have letters on one side and numbers on the other. This rule they are supposed to take on trust. Then they are given another rule by the same information source and they are supposed *not* to trust it but seek evidence for its falsity. If they do not continue to trust the first rule, then their card selections should diverge from Wason's expectations. If they simply forget about the background rule, the proper card choice would be A,K and 7; and if they want to test the background rule as well as the foreground rule, they would have to turn *all* cards. Notice that with the deontic interpretation, this split communication situation does not arise. The law stands and the task is to decide whether some people other than the source obey it. Here is an example of a subject who takes both rules on trust:

Subject 3. [Standard Wason task; has chosen A and 4]

- E. Why pick those cards and not the other cards?
- S. Because they are mentioned in the rule and I am assuming that the rule is true.

Another subject was rather bewildered when upon turning A he found a 7:

Subject 8.

- S. Well there is something in the syntax with which I am not clear because it does not say that there is an exclusion of one thing, it says 'if there is an A on one side there is a 4 on the other side'. So the rule is wrong.
- E. This [pointing to A] shows that the rule is wrong.
- S. Oh so the rule is wrong, it's not something I am missing.

Although this may sound similar to Wason's 'verification bias', it is actually very different. Wason assumed that subjects would be in genuine doubt about the truth value of the rule, but would then proceed in an 'irrational', verificationist manner to resolve the issue. What transpires here is that subjects take it on the authority of the experimenter that the rule is true, and then interprets the instructions as indicating those cards which are evidence of this:

Subject 22.

- S. Well my immediate [inaudible] first time was to assume that this is a true statement, therefore you only want to turn over the card that you think will satisfy the statement.

The communicative situation of the two-rule task is already much less bizarre, since there is no longer an reason to doubt the veracity of the experimenter. The excerpts also suggest that a modified standard task in which the rule is attributed not to the experimenter but to an unreliable source, might increase the number of competence responses. It hardly needs emphasising anymore that these problems cannot arise in the case of a deontic rule.

2.4. Subjects' understanding of propositional connectives

As mentioned before, the tutorial dialogues were preceded by a paraphrase task, in which subjects were asked whether a statement involving a conditional is equivalent to a statement involving other logical connectives. A further striking observations from the paraphrase task is that a conditional $p \rightarrow q$ is often (>50%) interpreted as a conjunction $p \wedge q$. Here is an example of what a conjunctive reading means in practice.

Subject 22. [Subject has chosen the conjunctive reading in the paraphrase task.]

- E. [Asks subject to turn the 7]
- S. That one . . . that isn't true. There isn't an A on the front and a 4 on the back. [. . .] you turn over those two [A and 4] to see if they satisfy it, because you already know that those two [K and 7] don't satisfy the statement.
- E. [baffled] Sorry, which two don't satisfy the rule?
- S. These two don't [K and 7], because on one side there is K and that should have been A, and that [7] wouldn't have a 4, and that wouldn't satisfy the statement.
- E. Yes, so what does that mean . . . you didn't turn it because you thought that it will not satisfy?
- S. Yes.

Clearly, on a conjunctive reading, the rule is already falsified by the cards as exhibited; no turning is necessary. The subject might however feel forced by the experimental situation to select some cards, and accordingly reinterprets the task as *checking* whether a given card satisfies the rule. This brings us to an important consideration: how much of the problem is caused by the conditional?

The literature on the selection task, with very few exceptions, has assumed that the problem is a problem specific to conditional rules. Indeed, it would be easy to infer also from the foregoing discussion of descriptive conditional semantics that the conditional (and its various expressions) is unique in causing subjects so much difficulty in the selection task, and that our only point is that a sufficiently rich range of interpretations for the conditional must be used to frame psychological theories of the selection task.

However, the issues already discussed—the nature of truth, response to exceptions, contingency, pragmatics—are all rather general in their implications for the task of seeking evidence for truth. One can distinguish the assessment of truth of a sentence from truthfulness of an utterer for sentences of any form. The robustness or brittleness of statements to counterexamples is an issue which arises for any generalisation. The social psychological effects of the experimenter's authority, and the communicative complexities introduced by having to take a cooperative stance toward some utterances and an adversarial one toward others is also a general problem of pragmatics that can affect statements of any logical form (cf. Grice, 1975). Contingencies between feedback from early evidence on choice of subsequent optimal evidence seeking are general to any form of sentence for which more than one case is relevant.

It would seem to be a high priority to find out to what extent there is something uniquely problematic about conditionals in the selection task, and to what extent these more general issues could explain poor performance in seeking evidence for descriptive statements' truth. Several early papers compared disjunction with the conditional (e.g., van Duyne, 1974), and

show that disjunctions are at least as hard as conditionals in the selection task. It is true that Johnson-Laird and Byrne (2002) cite Wason and Johnson-Laird (1969) as showing that disjunctions are easy, but that paper has so many divergences from the standard selection task it is hard to know how to relate it. But disjunction is perhaps too closely allied to the conditional to make a case that the problem is a more general problem of the possibility of non-truth-functionality of all natural language connectives. What would happen, for example, if the rule were stated using the putatively least problematical connective, conjunction?

2.5. Other sources of difficulty

The transcripts of the tutorial dialogues reveal another important source of confusion, namely the interpretation of the anaphoric expression 'one side . . . other side' and its interaction with the direction of the conditional. The trouble with 'one side . . . other side' is that in order to determine the referent of 'other side', one must have kept in memory the referent of 'one side'. That may seem harmless enough, but in combination with the various other problems identified here, it may prove too much. Even apart from limitations of working memory, subjects may have a non-intended interpretation of 'one side . . . other side', wherein 'one side' is interpreted as '*visible* side' (the front, or face of the card) and 'other side' is interpreted as '*invisible* side' (the back of the card). The expression 'one side . . . other side' is then interpreted as deictic, not as anaphoric. This possibility was investigated by Gebauer and Laming (1997), who argue that deictic interpretation of 'one side . . . other side' and a biconditional interpretation of the conditional, both singly and in combination, are prevalent, persistently held, and consistently reasoned with. Gebauer and Laming present the four cards of the standard task six times to each subject, pausing to actually turn cards which the subject selects, and to consider their reaction to what is found on the back. Their results show few explicitly acknowledged changes of choice, and few selections which reflect implicit changes. Subjects choose the same cards from the sixth set as they do from the first. Gebauer and Laming argue that the vast majority of the choices accord with normative reasoning from one of the four combinations of interpretation achieved by permuting the conditional/biconditional with the deictic/anaphoric interpretations.[11]

We tried to find further evidence for Gebauer and Laming's view, and presented subjects with rules in which the various possible interpretations of 'one side . . . other side' were spelt out explicitly; e.g. one rule was

(1) if there is a vowel on the face of the card, then there is an even number on the back
     To our surprise, subjects seemed completely insensitive to the wording of the rule and chose according to the standard pattern whatever the formulation; for discussion see Stenning and van Lambalgen (2001).
     This result made us curious to see what would happen in tutorial dialogues when subjects are presented with a rule like (1), and indeed the slightly pathological (2)
(2) if there is a vowel on the back of the card, there is an even number on the face of the card.

After having presented the subjects with these two rules, we told them that the *intended* interpretation of 'one side . . . other side' is that 'one side' can refer to the visible face or to the invisible back. Accordingly, they now had choose cards corresponding to

(3) if there is a vowel on one side (face or back), then there is an even number on the other side (face or back).

We now provide a number of examples, culled from the tutorial dialogues, which demonstrate the interplay between the interpretations chosen for anaphora and conditional. The first example shows us a subject who explicitly changes the direction of the implication when considering the back/face anaphora, even though she is at first very well aware that the rule is not biconditional.

Subject 12. [Experiments (1), (2), and (3)]

E. The first rule says that if there is a vowel on the face of the card, so what we mean by face is the bit you can see, then there is an even number on the back of the card, so that's the bit you can't see. So which cards would you turn over to check the rule.

S. Well, I just thought 4, but then it doesn't necessarily say that if there is a 4 that there is a vowel underneath. So the A.

E. For this one it's the reverse, so it says if there is a vowel on the back, so the bit you can't see, there is an even number on the face; so in this sense which ones would you pick?

S. [Subject ticks 4] This one.

E. So why wouldn't you pick any of the other cards?

S. Because it says that if there is an even number on the face, then there is a vowel, so it would have to be one of those [referring to the numbers].

⋮

E. [This rule] says that if there is a vowel on one side of the card, either face or back, then there is an even number on the other side, either face or back.

S. I would pick that one [the A] and that one [the 4].

E. So why?

S. Because it would show me that if I turned that [pointing to the 4] over and there was an A then the 4 is true, so I would turn it over. Oh, I don't know. This is confusing me now because I know it goes only one way.

⋮

S. No, I got it wrong didn't I, it is one way, so it's not necessarily that if there is an even number then there is a vowel.

The second example is of a subject who gives the normative response in experiment 3, but nonetheless goes astray when forced to consider the back/face interpretation.

Subject 4. [Experiments (1), (2), and (3)]

E. OK This says that if there is a vowel on the face [pointing to the face] of the card, then there is an even number on the back of the card. How is that different to . . .

S. Yes, it's different because the sides are unidirectional.

S. So would you pick different cards?

S. If there is a vowel on the face . . . I think I would pick the A.

E. And for this one? [referring to the second statement] This is different again because it says if there is a vowel on the back . . .

S. [completes sentence] then there is an even number on the face. I think I need to turn over the 4 and the 7. Just to see if it (the 4) has an A on the back.

E. OK Why wouldn't you pick the rest of the cards?

S. I'm not sure, I haven't made up my mind yet. This one (the A) I don't have to turn over because it's not a vowel on the back, and the K is going to have a number on the back so that's irrelevant. This one [the 4] has to have a vowel on the back otherwise the rule is untrue. I still haven't made up my mind about this one (the 7). Yes, I do have to turn it over because if it has a vowel on the back then it would make the rule untrue. So I think I will turn it over. I could be wrong.

[When presented with the rule where the anaphora have the intended interpretation]

S. I would turn over this one (the A) to see if there is an even number on the back and this one (the 7) to see if there was a vowel on the back.

Our third example is of a subject who explicitly states that the meaning of the implication must change when considering back/face anaphora.

Subject 16. [Experiments (1), (2), and (3)] [Subject has correctly chosen A in condition (1).]

E. The next one says that if there is a vowel on the back of the card, so that's the bit you can't see, then there is an even number on the face of the card, so that's the bit you can see; so that again is slightly different, the reverse, so what would you do?

S. Again I'd turn the 4 so that would be proof but not ultimate proof but some proof . . .

E. With a similar reasoning as before?

S. Yes, I'm pretty sure what you are after . . . I think it is a bit more complicated this time, with the vowel on the back of the card and the even number, that suggests that if and only if there is an even number there can be a vowel, I think I'd turn others just to see if there was a vowel, so I think I'd turn the 7 as well.

[In condition 3 chooses A and 4]

We thus see that, in these subjects, the direction of the conditional is related to the particular kind of deixis assumed for 'one side . . . other side'. This shows that the process of natural language interpretation in this task need not be compositional, and that, contrary to Gebauer and Laming's claim, subjects need not have a persistent interpretation of the conditional.

Two questions immediately arise:

1. why would there be this particular interaction?
2. what does the observed interaction tell us about performance in the standard Wason task?

Question 2 can easily be answered. *If* subjects would decompose the anaphoric expression 'one side . . . other side' into two deictic expressions 'face/back' and 'back/face' and would then proceed to reverse the direction of the implication in the latter case, they should choose the p and q cards. Also, since the expression 'one side . . . other side' does not appear in a deontic rule such as 'if you want to drink alcohol, you have to be over 18', subjects will not be distracted by this particular difficulty.

Question 1 is not answered as easily. There may be something pragmatically peculiar about a conditional of which the consequent, but not the antecedent, is known. These are

often used for diagnostic purposes (also called *abduction*): if we have a rule which says 'if switch 1 is down, the light is on', and we observe that the light is on, we are tempted to conclude that switch 1 must be down. This however is making an inference, not stating a conditional; but then subjects are perhaps not aware of the logical distinction between the two.

It is of interest that the difficulty discussed here was already identified by Wason and Green (1984), albeit in slightly different terms: their focus is on the distinction between a *unified* and a *disjoint* representation of the stimulus. A unified stimulus is one in which the terms referred to in the conditional cohere in some way (say as properties of the same object, or as figure and ground), whereas in a disjoint stimulus the terms may be properties of different objects, spatially separated.

Wason and Green conjectured that it is disjoint representation which accounts for the difficulty in the selection task. To test the conjecture they conducted three experiments, varying the type of unified representation. Although they use a reduced array selection task (RAST), in which one chooses only between q and $\neg q$, relative performance across their conditions can still be compared.

Their contrasting sentence rule pairs are of great interest, partly because they happen to contain comparisons of rules with and without anaphora. There are three relevant experiments numbered 2–4. Experiment 2 contrasts unified and disjoint representations without anaphora in either, and finds that unified rules are easier. Experiment 3 contrasts unified and disjoint representations with the disjoint rule having anaphora. Experiment 4 contrasts unified and disjoint representations but removes the anaphora from the disjoint rule while adding another source of linguistic complexity (an extra tensed verb plus pronominal anaphora) to the unified one. For a full discussion of their experiments we refer the reader to Stenning and van Lambalgen (2001)Stenning and van Lambalgen (2001); here we discuss only experiment 2.

In their experiment 2, cards show shapes (triangles, circles) and colours (black, white), and the two sentences considered are

  (4) Whenever they are triangles, they are on black cards.
  (5) (2b) Whenever there are triangles below the line, there is black above the line.

That is, in (4) the stimulus is taken to be unified because it is an instance of figure/ground, whereas in (5) the stimulus consists of two parts and hence is disjoint. Performance for sentence (4) was worse than for sentence (4) (for details see Wason & Green, 1984, pp. 604–607).

We would describe the situation slightly differently, in terms of the contrast between deixis and anaphora. Indeed, the experimental set-up is such that for sentence (5), the lower half of the cards is hidden by a bar, making it analogous to condition (2), where the object mentioned in the antecedent is hidden. We have seen above that some subjects have difficulties with the intended direction of the conditional in experiment (2). Sentence (5) would be the 'difficult half' of the anaphora-containing sentence "Whenever there are triangles on one side of the line, there is black on the other side of the line". Sentence (4) does not contain any such anaphora. With Wason and Green we would therefore predict that subjects find (5) more difficult.

3. Experiment

In this experiment, several conditions are compared with base-line performance on the classical descriptive 'abstract' task, each designed to assess the contribution to determination of choice by one of the factors discussed above. We describe each condition in turn, and then present the results together.

3.1. The conditions

3.1.1. Classical 'abstract' task
To provide a baseline of performance on the selection task with descriptive conditionals, the first condition repeats Wason's (1968) classical study with the following instructions and materials (see instructions in Section 1). The other conditions are described through their departures from this baseline condition.

3.1.2. Two-rule task
After the preliminary instructions for the classical task, the following instructions were substituted in this condition:

> . . . Also below there appear two rules. One rule is true of all the cards, the other isn't. Your task is to decide which cards (if any) you *must* turn in order to decide which rule holds. Don't turn unnecessary cards. Tick the cards you want to turn.

Rule 1: *If there is a vowel on one side, then there is an even number on the other side.*
Rule 2: *If there is a consonant on one side, then there is an even number on the other side.*
Normative performance in this task, according to the classical logical competence model, is to turn only the not-Q card. The rules are chosen so that the correct response is to turn exactly the card that the vast majority of subjects fail to turn in the classical task. This has the added bonus that it is no longer correct to turn the P card which provides an interesting comparison with the original task. This is the only descriptive task for which choosing the true-antecedent case is an error.

By any obvious measure of task complexity, this task is more complicated than the classical task. It demands that two conditionals are processed and that the implications of each case are considered with respect to both rules and with respect to a distribution of truth values. Nevertheless, our prediction was that performance should be substantially nearer the logically normative model for the reasons described above.

3.1.3. Contingency instructions
The 'contingency instructions', designed to remove any difficulties in understanding that choices have to be made ignoring possible interim feedback, after an identical preamble, read as follows, where the newly italicised portion is the change from the classical instructions:

> . . . Also below there appears a rule. Your task is to decide which of these four cards you *must* turn (if any) in order to decide if the rule is true. *Assume that you have to decide whether to turn each card before you get any information from any of the turns you choose to make.* Don't turn unnecessary cards. Tick the cards you want to turn.

If the contingencies introduced by the descriptive semantics are a source of difficulty for subjects, this additional instruction should make the task easier. In particular, since there is a tendency to choose the P card first, there should be an increase in not-Q responding.

After conducting this experiment we found a reference in Wason (1987) to use of essentially similar instructions in his contribution to the Science Museum exhibition of 1977, and there are mentions in other early papers. Clearly he had thought about assumed contingencies between card choices as a possible confusion. Wason reports no enhancement in his subjects' reasoning, but he does not report whether any systematic comparison between these and standard instructions was made, or quite what the population of subjects were.

3.1.4. Judging truthfulness of an independent source

We chose to investigate the possible contribution of problems arising from the authoritative position of the experimenter and the balance of cooperative and adversarial stances required toward different parts of the task materials through instructions to assess truthfulness of the source instead of truth of the rule, and we separated the source of the rule from the source of the instructions (the experimenter). The instructions read as follows:

> ... Also below there appears a rule *put forward by an unreliable source*. Your task is to decide which cards (if any) you *must* turn in order to decide *if the unreliable source is lying*. Don't turn unnecessary cards. Tick the cards you want to turn.

With these instructions there should be no discomfort about seeking to falsify the rule. Nor should any falsity of the rule throw any doubt on the truthfulness of the rest of the instructions, since the information sources are independent.

These 'truthfulness' instructions are quite closely related to several other manipulations that have been tried in past experiments. In the early days of experimentation on this task, when it was assumed that a failure to try and falsify explained the correct response, various ways of emphasising falsification were explored. Wason (1968) instructed subjects to pick cards which could break the rule and Hughes (1966) asked them whether the rule was a lie. Neither instruction had much effect. However, these instructions fail to separate the source of the rule from the experimenter (as the utterer of the rule) and may fail for that reason.

Kirby (1994) used a related manipulation in which the utterer of the rule was a machine said to have broken down, needing to be tested to see if it was working properly again after repair. These instructions did produce significant improvement. Here the focus of the instruction is to tell whether the machine is 'broken', not simply whether the utterance of the rule is a falsehood. This might be expected to invoke a deontic interpretation (Kirby's condition is akin to the 'production line inspection scenarios' mentioned before), and so it might be that the improvement observed is for this reason.

Platt and Griggs (1993) explored a sequence of instructional manipulations in what they describe as abstract tasks which culminate in 81% correct responding. One of the changes they make is to use instructions to 'seek violations' of the rule, which is relevant here for its relation to instructions to test the truth of an unreliable source. Their experiments provide some insight into the conditions under which these instructions do and don't facilitate performance. Platt and Griggs study the effect of 'explications' of the rule and in the most effective manipulations actually replace the conditional rule by explications such as: 'A card with a vowel on it can only

have an even number, but a card with a consonant on it can have either an even or an odd number.' Note that this explication removes the problematic anaphora (see above, Section 2.5), explicitly contradicts a biconditional reading, and removes the conditional, with its tendency to robust interpretation. But more significantly still, the facilitation of turning not-Q is almost entirely effected by the addition of 'seek violations' instructions, and these instructions probably switch the task from a descriptive to a deontic task.

In reviewing earlier uses of the 'seek violations' instruction Platt and Griggs note that facilitation occurs with abstract permission and obligation rules but not with the standard abstract task. So, merely instructing to seek violations doesn't invoke a deontic reading when the rule is still indicative, and the instruction is still interpretable descriptively—'violations' presumably might make the rule false. But combined with an 'explication' about what cards *can* have on them (or with permission or obligation schema) they appear to invoke a deontic reading. As we shall see, 80% seems to be about the standard rate of correct responding in deontically interpreted tasks regardless of whether they contain material invoking social contracts.

So the present manipulation does not appear to have been explored before. We predicted that separating the source of the rule from the experimenter while maintaining a descriptive reading of the rule should increase normative responding.

3.1.5. Exploring other kinds of rules than conditionals

This condition of the experiment was designed to explore the malleability of subjects' interpretations of rules other than conditionals. In particular we chose a conjunctive rule as arguably the simplest connective to understand. As such this condition has a rather different status from the others in that it is not designed to remove a difficulty from a logically similar task but to explore a logical change. Since it was an exploration we additionally asked for subjects' justification of their choices afterwards.

A conjunctive rule was combined with the same instructions as are used in the classical abstract task.

Rule: *There is a vowel on one side, and there is an even number on the other side.*

The classical logical competence model demands that subjects should turn no cards with such a conjunctive rule—the rule interpreted in the same logic as Wason's interpretation of his conditional rule can already be seen to be false of the not-P and not-Q cards. Therefore, under this interpretation the rule is already known to be false and no cards should be turned.

We predicted that many subjects would not make this interpretation of this response. An alternative, perfectly rational, interpretation of the experimenter's intentions is to construe the rule as having deontic force (every card *should* have a vowel on one side and an even number on the other) and to seek cards which might flout this rule other than ones that obviously can already be seen to flout it. If this interpretation were adopted, then the P and Q cards would be chosen. Note that this interpretation is deontic even though the rule is syntactically indicative.

3.2. Subjects

Subjects were 377 first year Edinburgh undergraduates, from a wide range of subject backgrounds.

Table 1
Frequencies of card choice combinations by conditions

Condition	P Q	Q	P	P ¬Q	¬Q	¬P,Q	P,Q, ¬Q	¬P,¬Q	All	None	Miscellaneous	Total
Classical	56	7	8	4*	3	7	1	2	9	8	5	108
Two-rule	8	8	2	1	9*	2	1	0	0	2	4	37
Contingency	15	0	3	8*	1	6	4	8	3	0	3	51
Truthfulness	39	6	9	14*	0	7	3	6	8	15	5	112
Conjunction	31	2	9	7	2	0	0	1	0	9*	8	69

Classical logical competence responses are marked asterisk (*). Any response made by at least three subjects in at least one condition is categorised: everything else is miscellaneous.

3.3. Method

All tasks were administered to subjects in classroom settings in two large lectures. Subjects were randomly assigned to the different conditions, with the size of sample in each condition being estimated from piloting on effect sizes. Adjacent subjects did different conditions. The materials described above were preceded by the following general instruction:

> The following experiment is part of a program of research into how people reason. Please read the instructions carefully. We are grateful for your help.

3.4. Results

Those subjects (12 across all conditions) who claimed to have done similar tasks before, or to have received any instruction in logic were excluded from the analysis.

Table 1 presents the data from all of the conditions. Any response made by at least three subjects in at least one condition is categorised: all other responses are treated as miscellaneous. Subjects were scored as making a completely correct response, or as making at least some mistake, according to the classical logical competence model. For all the conditions except the two-rule task and the conjunction condition, this 'competence model' performance is choice of P and not-Q cards. For the two-rule task the correct response is not-Q. For the conjunction condition it is to turn no cards.

Table 2 presents the tests of significance of the percentages of correct/incorrect responses as compared to the baseline classical condition. 3.7% of subjects in the baseline condition made the correct choice of cards.

The percentages completely correct in the other conditions were two-rule condition 24%; 'truthfulness' condition 13%; in the 'contingency' condition 18%; and in the conjunction condition 13%. The significance levels of these proportions by Fisher's exact test appear in Table 2.

The two-rule task elicits substantially more competence model selections than the baseline task. In fact the completely correct response is the modal response. More than six times as many subjects get it completely correct even though superficially it appears a more complicated task. The next most common responses are to turn P with Q, and to turn just Q. The former is the modal response in the classical task. The latter appears to show that even with unsuccessful

Table 2
Proportions of subjects completely correct and significances of differences from baseline of each of the four manipulations

Condition	Wrong	Right	P	Percent correct
Classical baseline	104	4		3.7
Two-rule	28	9	0.004	24
Contingency	37	8	0.005	18
Truthfulness	98	14	0.033	13
Conjunction	60	9	0.022	13

subjects, this task shifts attention to the consequent cards—turnings of P are substantially suppressed: 32% as compared to 80% in the baseline task.

Contingency instructions also substantially increase completely correct responding, and do so primarily at the expense of the modal P with Q response. In particular they increase not-Q choice to 50%.

Instructions to test the truthfulness of an unreliable source have a smaller effect which takes a larger sample to demonstrate, but nevertheless, 13% of subjects get it completely correct, nearly four times as many as the baseline task. The main change is again a reduction of P with Q responses, but there is also an increase in the response of turning nothing.

Completely correct performance with a conjunctive rule was 13%—not as different from the conditions with conditional rules as one might expect if conditionals are the main source of difficulty. The modal response is to turn the P and Q cards—just as in the original task. Anecdotally, debriefing subjects after the experiment reveals that a substantial number of these modal responses are explained by the subjects in terms construable as a deontic interpretation of the rule, roughly paraphrased as "The cards should have a vowel on one side and and even number on the other". The P-with-Q response is correct for this interpretation.

3.5. Discussion of results

Each of the manipulations designed to facilitate reasoning in the classical descriptive task makes it substantially easier as predicted by the semantic/pragmatic theories that the manipulations were derived from. The fact that subjects' reasoning is improved by each of these manipulations, provides strong evidence that subjects' mental processes are operating with related categories in the standard laboratory task. Approaches like those of Sperber's Relevance Theory propose that the subjects solve the task 'without thinking'. The fact that these instructional manipulations have an impact on subjects' response strongly suggests that the processes they impact on are of a kind to interact with the content of the manipulations. This still leaves the question at what level of awareness? But even here, the tutorial dialogues suggest that the level is not so far below the surface as to prevent these processes being quite easily brought to some level of awareness.

It is important to resist the idea that if subjects were aware of these problems, that itself would lead to their resolution, and the conclusion that therefore subjects can't be suffering these problems. Extensive tutoring in the standard task which is sufficient to lead subjects

to make their problems quite explicit, generally does *not* lead, at least immediately, to stable insight. This is as we should expect. If, for example, subjects become aware that robustness to counterexamples makes the task instructions uninterpretable, that itself does not solve their problem of how to respond. Or, for another example, if subjects become aware of being unable to reflect contingencies between choices in their responses, that does not solve the problem of what response to make. General questions of what concepts subjects have for expressing their difficulties, and in what ways they are aware of them are important questions, especially for teaching. These questions invite further research through tutoring experiments, but they should not be allowed to lead to misinterpretation of the implications of the present results. We take each condition in turn

3.5.1. The two-rule task

There are other possible explanations as to how the novel task functions to facilitate competence model responding. If subjects tend to confuse the two situations: "this rule is true of this card" and "this card makes this rule true" then it may help them that the two-rule task is calculated to lead them early to a conflict that a single card (e.g., the true consequent card) "makes both rules true" even as the instructions insist that one rule is true and one false. Although some subjects may infer that there must therefore be something wrong with the instructions, others progress from this impasse to appreciate that cases can comply with a rule without making it true—the semantic relations are asymmetrical even though the same word 'true' can, on occasion, be used for both directions. This confusion between semantic relations is evidently closely related to what Wason early called a 'verification' strategy (searching for compliant examples) in that it may lead to the same selections, but it is not the strategy as understood by Wason. This confusion between semantic relations is in abundant evidence in the dialogues.

The two-rule task makes an interesting comparison with at least three other findings in the literature. First, the task was designed partly to make explicit the choice of hypotheses which subjects entertain for the kind of rational choice modelling proposed by Oakford and Chater. Providing two explicit rules (rather than a single rule to be compared with an assumed null hypothesis of independence) makes the false-consequent card unambiguously the most informative card and therefore the one which these models should predict will be most frequently chosen. In our data for this task, the false-consequent card comes in third substantially behind the true antecedent and true consequent cards.

For a second comparison, Gigerenzer and Hug (1992) studied a manipulation which is of interest because it involves both a change from deontic to descriptive interpretation and from single to two-rule task. One example scenario, had a single rule that hikers who stayed overnight in a hut had to bring their own firewood. Cards represented hikers or guides and bringers or non-bringers of wood. As a single-rule deontic task with instructions to see whether people obeyed the rule, this produced 90% correct responding, a typical result. But when the instructions asked the subject to turn cards in order to decide whether this rule was in force, or whether it was the guides who had to bring the wood, then performance dropped to 55% as conventionally scored. Gigerenzer and Hug explain this manipulation in terms of 'perspective change', but this is both a shift from a deontic task to a descriptive one (in the authors' own words 'to judge whether the rule is *descriptively* wrong' (our emphasis), and from a single rule to a two-rule task, albeit that the second rule is mentioned but not printed alongside its alternative.

Unfortunately, the data cannot be scored appropriately for the classical competence model for the two-rule task from what is presented in the paper, but it appears to produce a level of performance higher than single rule abstract tasks but lower than deontic tasks, just as we observe. Direct comparison of the two subject populations is difficult as Gigerenzer's subjects score considerably higher on all the reported tasks than ours, and no baseline single-rule descriptive task is included.

The third comparison of the two-rule task is with work on 'reasoning illusions' by Johnson-Laird and coworkers mentioned above (Johnson-Laird & Byrne, 2002; Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000; Johnson-Laird & Savary, 1999). Johnson-Laird and Savary (1999, p. 213) presented exactly comparable premises to those we used in our two-rule task but asked their subjects to choose a conclusion, rather than to seek evidence about which rule was true and which false. Their interest in these problems is that mental models theory assumes that subjects 'only represent explicitly what is true', and that this gives rise to 'illusory inferences'. The following material was presented with the preface that both statements are about a hand of cards, and one is true and one is false:

1. If there is a king in the hand, then there is an ace.
2. If there is *not* a king in the hand, then there is an ace.

Select one of these conclusions:

1. There is an ace in the hand
2. There is not an ace in the hand
3. There may or may not be an ace in the hand.

Johnson-Laird and Savary (1999) report that 15 out of 20 subjects concluded that that there *is* an ace in the hand, and the other five concluded that there might or might not be an ace in the hand. They claim that the 15 subjects are mistaken in their inference.

> Hence, apart from one caveat to which we will return, there is no reasonable interpretation of either the disjunction or the conditionals that yields a valid inference that there is an ace. (p. 204)

The caveat appears to be that there are interpretations on which the premises are inconsistent and therefore *anything* (classically) logically follows, including this conclusion (p. 220).

What struck us initially is that our subjects show some facility with reasoning about assumptions of the same form even when our task also requires added elements of selection rather than merely inference. Selection tasks are generally harder. Specifically, our two-rule task introduces the circumstance which Johnson-Laird and Savary claim mental models predicts to introduce fundamental difficulty, i.e. reasoning from knowledge that some as yet unidentifiable proposition is false. This introduction makes the selection task much *easier* for subjects than its standard form in our experiment.

On a little further consideration, there is at least one highly plausible interpretation which make this conclusion valid and is an interpretation which appears in our dialogues from the two-rule task. Subjects think in terms of one of the rules *applying* and the other not, and they confuse (not surprisingly) the semantics of applicability with the semantics of truth. This is exactly the semantics familiar from the *IF . . . THEN . . . ELSE* construct of imperative

computer languages. If one clause applies and the other doesn't then it follows that there is an ace. Whether the alternativeness of the rules is expressed metalinguistically (by saying one is false and one true) or object-linguistically (with an exclusive disjunction), and whether the rules are expressed as implications or as exclusive disjunctions, thinking in terms of applicability rather than truth is a great deal more natural and has the consequence observed. Johnson-Laird (personal communication) objects that this interpretation just is equivalent to the mental models theory one. But surely this is a crisp illustration of a difference between the theories. If an interpretation in terms of applicability is taken seriously, subjects *should* draw this conclusion, and should stick to it when challenged (as many do). In fact failure to draw the inference is an error under this interpretation. Only mental models theory's restriction to a range of classical logical interpretations makes it define the inference as an error. We will put our money on the subjects having the more plausible interpretation of the conditionals here and the experimenters suffering an illusion of an illusion.

3.5.2. Contingency instructions

As mentioned above, effects of this manipulation have been reported by Wason in early studies, but his theory of the task did not assign it any great importance, or lead him to systematically isolate the effect, or allow him to see the connection between descriptive interpretation and this instruction. In the context of our hypothesis that it is descriptive versus deontic interpretation which is the main factor controlling difficulty of the task through interactions between semantics and instructions, this observation that contingency has systematic and predicted effects provides an explanation for substantial differences between the abstract task and content facilitations which invoke deontic interpretations. None of the other extant theories assign any significant role to this observation.

The effectiveness of contingency instructions presents particular difficulties for current rational choice models, since the choice of false-consequent cards rises so dramatically with an instruction which should have no effect on the expected information gain.

3.5.3. Truthfulness instructions

As described above, the truthfulness condition differs from past attempts to cue subjects to seeking counterexamples. Its success in bringing about a significant if small improvement may have resulted from effects of the manipulation other than the social psychological effects or the more general pragmatic effects of the balance of cooperative and adversarial stances described above. For example, it may well be that at least some subjects are more adept at thinking about the truthfulness of speakers than the truth values of their utterances abstracted from such issues as ignorance or intent to deceive.

3.5.4. The conjunctive rule

The purpose behind the conjunctive version of the task was rather different from the other manipulations, namely to show that many features of the task militate against the adoption of Wason's intended interpretation of his instructions quite apart from difficulties specific to conditionals. The interpretation of sentence semantics is highly malleable under the forces of task pragmatics. The results show that a conjunctive rule is treated very like (even if significantly differently from) the *if . . . then* rule. A higher proportion of subjects make the 'classically

correct' response than in the baseline task (13% as compared to 3.7%) but the modal response is the same (P and Q) and is made by similar proportions of subjects (45% conjunctive as compared to 52% baseline). One possibility is that a substantial number of subjects adopt a deontic interpretation of the rule and are checking for the cards that might be violators but are not yet known to be.

It is also possible that these results have more specific consequences for interpretation of the standard descriptive task. We know from Fillenbaum's (1978) work and from our own paraphrase tasks (Stenning & van Lambalgen, 2001) that about a half of subjects most readily entertain a conjunctive reading of *if . . . then* sentences. The developmental literature reviewed in Evans, Newstead, and Byrne (1993) reveals this interpretation to be even commoner amongst young children. It is most implausible that this interpretation is due merely to some polysemy of the connective 'if . . . then'. Much more plausible is that the conjunctive reading is the result of assuming the truth of the antecedent suppositionally, and then answering subsequent questions from within this suppositional context.

Be that as it may, if subjects' selections in the conditional rule tasks correspond to the selections they would make given an explicit conjunction in the conjunction condition, and we are right that these selections are driven in this condition by an implicitly deontic interpretation of the conjunction, then this suggests a quite novel explanation of at least some 'matching' responses in the original conditional task. Perhaps the similar rate of choice of P and Q in the conjunction and 'if . . . then' conditions points to a substantial number of subjects applying a deontic conjunctive interpretation in the standard task?

This hypothesis in turn raises the question how such a reading would interact with negations in the 'negations' paradigm which is the source of the evidence for Evan's (1972) 'matching' theory and therefore the source of one leg of 'dual process' theory (Evans & Over, 1996)? If interpretations stemming from deontic readings tend strongly toward wide sentential scope for negation, then one would predict that the rule with negated antecedent would be read as 'Its not the case that there is a vowel on one side and an even number on the other' which would lead to the same choices of A and 4, though for opposite reasons. That is, K and the 7 are now seen as already compliant, and the A and the 4 have to be tested to make sure they *don't* have an even number or a vowel, respectively. Pursuing this line of thought further suggests that negations in the second clause may not be interpretable in this framework (because of their interactions with the anaphors) and subjects might be forced to interpret them with the same wide scope, again leading to the same card choices, and potentially explaining why 'matching' appears to be unaffected by negation. Providing a semantic explanation, of course leaves open the questions about what processes operate. Evidently, further research will be required to explore these possibilities. The semantic analyses may seem complex but they make some rather strong predictions about how subjects should react to card turnings. This is an interesting line for future research holding out the possibility of a semantic basis for matching behaviour.

One objection to these various interpretations of the conjunction condition results might be that there are other interpretations of the rule used. Subjects might, for example, have interpreted the rule existentially, as claiming that at least one card had a vowel on one side and an even number on the other. This would lead normatively to the same A and 4 selections.

Accordingly, in a follow-up experiment, we revised the conjunctive rule to:

Rule: *There are vowels on one side of the cards and even numbers on the other.*

Table 3
Frequencies of card choice combinations by conditions

Condition	P Q	Q	P	P ¬Q	¬Q	¬P,Q	P,Q, ¬Q	¬P,¬Q	All	None	Miscellaneous	Total
Classical	56	7	8	4*	3	7	1	2	9	8	5	108
Two-rule	8	8	2	1	9*	2	1	0	0	2	4	37
Contingency	15	0	3	8*	1	6	4	8	3	0	3	51
Truthfulness	39	6	9	14*	0	7	3	6	8	15	5	112
Conjunction	31	2	9	7	2	0	0	1	0	9*	8	69
Baseline 2	10	2	10	1*	1	0	0	0	0	1	3	30
Conjunction 2	21	1	3	1	0	0	0	0	0	1*	2	30
Abstract subjunctive	13	2	8	3*	0	1	2	1	1	0	0	31

The modified conjunction task and its new baseline condition are below the earlier results which are repeated here for convenience. Classical logical competence responses are marked asterisk (*). Any response made by at least three subjects in at least one condition is categorised: everything else is miscellaneous.

It is implausible that this rule might be interpreted existentially. We ran this rule in another condition with its own baseline condition to ensure comparability of the new population. Table 3 shows the results of this experiment, with the earlier results repeated for convenient comparison. The result was slightly more extreme with this version of the conjunctive rule. 70% of subjects (rather than 45%) chose the P and Q cards. The proportion of classical logical competence model responses was identical to that for the baseline conditional task, and the baseline condition showed the population was comparable. The rewording raised the proportion of subjects giving the modal P and Q response. This rewording of the conjunctive rule appeared to make the universal deontic reading even less ambiguously the dominant reading.

These conjunctive rule results illustrate several general issues: how easy it is to invoke a deontic reading of indicative wording; how unnatural it is for naive subjects to adopt an 'is-this-sentence-literally-true' perspective rather than a 'what-are-the-experimenter's-intentions' perspective; that the difficulty of classical interpretation can be as great with conjunction as with implication. Although the difficulties may be different difficulties, there is a real possibility that they are closely related through conjunctive suppositional interpretations of the conditional.

Finally, we explored one other obvious manipulation designed to follow up the malleability of subjects' interpretations exposed by the conjunctive rule. If subjects' difficulties in the original descriptive task follow from the complexities of descriptive semantics, is it possible to restore deontic levels of performance in the abstract task merely by making the rule subjunctive? We ran a further condition in which the rule used was:

If a card has a vowel on one side, then it *should* have an even number on the other.

and the instruction was to choose which of the four cards you *must* turn in order to decide if the card complies with the rule.

The results of this condition are shown in Table 3 in the 'Abstract subjunctive' row. Three subjects of 31 turned P and not-Q, as compared to one of 30 in the baseline. If this is a facilitation it is a small one. Merely using subjunctive wording may be insufficient to invoke a deontic

reading. This is not so surprising since there is an alternative 'epistemic' interpretation of the subjunctive modal here which might still be used with a descriptive semantics for the underlying rule. Imagine that the rule is clearly a robust descriptive scientific law (perhaps 'All ravens are black'), then one might easily state in this context, that a card with 'raven' on one side *should* have 'black' on the other, implying something about what the cards have to be like to comply with the scientific law (still with a descriptive semantics underlying), rather than what the birds have to do to comply with a legal regulation. This possibility of interpretation may make it hard to invoke a deontic interpretation without further contentful support. Contentful support is, of course, what the various 'quality inspector' scenarios provide. Contentful support is also what permission and obligation schemas, and the 'seek violations' instructions in combination with modal explications of the rule provide, as reported by Platt and Griggs (1993).

In summary of all the conditions, these results corroborate the findings of the tutoring experiments, also reported in Stenning and van Lambalgen (2001), that our manipulations alleviate real sources of difficulty with interpretation for subjects in the original descriptive task—sources of difficulty which do not apply in the deontic task. This evidence suggests that far from failing to think at all, subjects are sensitive to several important semantic issues posed by the descriptive task.

4. General discussion

What implications do these results have for theories of reasoning, and for the place of interpretation in cognitive theory more generally?What do they tell us about the way the field has viewed the relation between logical and psychological analyses of reasoning, and how that relation might be construed more productively? Each theory is a somewhat different case.

These results remove the founding evidence for 'evolutionary' theories which propose that the difference in performance on 'social contract' conditionals and descriptive conditionals needs to be explained by innate cheating detection modules evolved in the Pleistocene. Our evidence is that the descriptive and deontic tasks are quite different tasks and that the former is fraught with interpretational problems where the latter is straightforward. So the selection task evidence has no direct bearing on innateness, modularity, or the Pleistocene, though it can be used to formulate some interesting and contrary hypotheses about cheating detection (Stenning, 2002).

More generally, this reappraisal of the selection task provides a good example of how arguments for 'massive modularity' in cognition should be treated with some scepticism. The original experiments found variation in performance as a function of difference in materials. Sweeping generalisations were then made from the laboratory task without any consideration of the relation between that task and subjects other communication and reasoning abilities. Just as our analysis directs attention to the differences between variations on the selection task and the continuities between natural language communication inside and outside the selection task, so our proposals return attention to the evolutionary issue how humans' generalised communication capacities arose in evolution. The interactions between logic's dual apparatus of interpretation and of derivation constitute an exquisitely context sensitive conceptual framework for the study of human reasoning and communication, whether in evolution, development or education.

The non-evolutionary theories of human reasoning are most generally affected by the present results through their implications for the relation between logic and psychology. We focus here particularly on relevance theory, mental models theory, and rational analysis models.

Inasmuch as relevance theory assumes that human reasoning and communication abilities are general abilities which interact with contextual specificities, our general drift is sympathetic to relevance theory's conclusions. We agree with relevance theory that the goal must be to make sense of what subjects are doing in the very strange situation of laboratory reasoning tasks—in a memorable phrase, to see subjects as 'pragmatic virtuosos' (Girotto, Kemmelmeier, Sperber, & van der Henst, 2001) rather than to see them as logical defectives. Our divergences from relevance theory are about the granularity of interaction between semantic and pragmatic processes in subjects' reasoning; in the range of behaviour we believe to be of theoretical concern; and in the program of research.

Relevance theory explains pragmatic effects in terms of very general factors—relevance to the task at hand and cost of inference to reveal that relevance. These factors must always operate with regard to some semantic characterisation of the language processed. Condensing analysis into these two pragmatic factors however seems, in this case at least, to have led to relevance theorists missing the critical *semantic* differences which drive the psychological processes in this task—the differences between deontic and descriptives and their consequences for interpretation in this task's setting. Relevance theory's conclusion has been that not much reasoning goes on when undergraduate subjects get the abstract task 'wrong'. Our combination of tutoring observations and experiment strongly suggest that a great deal goes on, however speedily the 'precomputed' attitudes are brought to bear in the actual task, and that the exact nature of the processes is highly variable from subject to subject. Taking logic more seriously leads us to seek more detailed accounts of mental processes.

The current results have rather wide-ranging implications for mental models theory. Some implications specific to the theory's application to the selection task have already been discussed. Others are more general, about mental models theory's relation to logic and semantics. Since Johnson-Laird's early work with Wason on the selection task mental models theory has been elaborated by a complex theory of the meanings of conditionals and the overlay of semantics by 'pragmatic modulation', and the theory has been much exercised by the issue whether subjects' interpretations of the rule in the selection task is truth-functional or not. However, this consideration of semantic possibilities has been divorced from any consideration of their implications for the subjects' interpretation of the *task*. If subjects' reading of the rule is non-truth-functional (by whatever semantic or pragmatic route), then the subject should experience a *conflict* between their interpretation and the task instructions. This conflict has never been acknowledged by mental models theorists. What justification can there then be for applying the classical logical competence model as a criterion of correct performance while simultaneously rejecting it as an account of how subjects interpret the conditional?

But the most significant implications of our analysis for mental models theory are implications for its general understanding of the relation between logic and psychology. Mental models theory and its opponents such as the 'mental logics' of Rips (1994), agree in assigning greatest prominence to the issue whether subjects reason using models or rules. Our claim is that both camps' interpretations of these logical concepts are too mechanical, and the consequence is that the psychological investigations fail to give empirical content to the distinction.

Modern logic formalises its concept of interpretation in model theory, and of derivation (including rules of inference) in proof theory. Of course, one can reason over types of model but only within some meta-language (often in practice a natural language such as English or a formal language such as set theory), which, of course, in turn requires its own proof-theory and rules of inference. So rules are involved in this reasoning too. One sees this issue exemplified in mental models theory, which of necessity must also include principles for the manipulation of models. The principles are rules for the manipulation of model representations, and are just as formal, linguistically specified and content free as rules of inference in sentential systems. In fact there are point-by-point correspondences, not just general equivalences. For the systems of concern in the psychology of reasoning, logic provides completeness proofs. The import of those proofs is that any inference described in a semantic way using models can be captured by a syntactic process using sentential rules. In fact models in mental models theory are what proof theorists call cases in a proof-by-cases strategy. Looking from the outside, as psychological researchers are forced to do, one cannot distinguish rules from models on the basis of observing merely the inputs and outputs of reasoning processes (see Hodges, 1993 for a logician's appraisal of mental models theory's account of its relation to logic). Coopting the interpretational apparatus of logic as a mechanism for modelling derivational processes, merely obscures the crucial distinction between interpretation and derivation.

These are general logical arguments about correspondences between classes of system. Stenning and Oberlander (1995) and Stenning and Yule (1997) provided detailed studies of the two most relevant particular equivalences between model and rule systems: mental models and Euler diagrams, and between mental models and a fragment of propositional logic. Stenning and van Lambalgen (submitted) provide a non-monotonic model of conditionals which shows how sentences and models work together in the processes of interpretation and reasoning. These arguments show that the issue of rules versus models has not yet been given any empirical content. The psychological debate misconstrues logic by treating it as providing mechanisms of reasoning, whereas it should be construed at a more abstract level. For example, our present proposals about the selection task claim that the dominant factor in determining reasoning will be whether subjects assign descriptive or deontic *form* to the rule presented. The processes of reasoning from either of these assigned interpretations can be formulated as sentential reasoning or as model-based reasoning, or as some combination of the two (see, e.g., Stenning & van Lambalgen, submitted for a treatment of the 'suppression' task in these terms).

Finally, where do our findings leave the rational analysis models of selection task behaviour as optimal experiment (Oaksford & Chater, 1994). We applaud these authors' challenge to the uniqueness of the classical logical model of the task, and also their insistence that the deontic and descriptive versions of the task require distinct accounts. This theory is clearly more sophisticated about the relations between formal models and cognitive processes than the theories it challenges. However, our proposals are quite divergent in their cognitive consequences. The rational analysis models reject any role for logic, claiming that the task is an inductive one. But this move smuggles logic in the back door. Applying optimal experiment theory requires assigning probabilities to propositions, and propositions are specified in some underlying language. The logic underlying the rational analysis model is the same old classical propositional calculus with all its attendant divergences from subjects' interpretations of the task materials. This has direct psychological consequences. The rational analysis models treat

subjects' performances as being equally correct as measured by the two distinct competence models for descriptive and deontic tasks. Our analysis predicts that the descriptive task will be highly problematical and the deontic task rather straightforward. The tutorial evidence on the descriptive task and its experimental corroboration support our prediction about the descriptive task. Approaching through interpretation predicts and observes considerable variety in the problems different subjects exhibit in the descriptive task, and even variety within the same subject at different times. We can agree that some subjects may adopt something like the rational analysis model of the task, but disagree about the uniformity of this or any other interpretation. Most of all we do not accept that everyone is doing the same thing at the relevant level of detail.

This situates our approach with regard to some prominent psychological theories of reasoning, and illustrates similarities and differences with extant approaches in the context of this one particular task. But our proposals also have general implications for how cognitive theories of reasoning relate to logical and linguistic theories of language and communication more generally. If we are anything like right about the selection task, it is both possible and necessary to bring the details of formal accounts of natural languages (semantics of deontics and descriptives, variable and constant anaphora, tense, definiteness, domain of interpretation, scope of negation, . . .) to bear in explaining the details of performance in laboratory reasoning tasks. This is necessary because subjects' behaviour in these tasks is continuous with generalised human capacities for communication, and possible because although strange in many ways, laboratory tasks have to be construed by subjects using their customary communicative skills. Once this apparatus is transferred to the psychological laboratory, it can yield powerful explanatory theories of why small details of the materials yield large changes in behaviour. For example, the empirical evidence is that the dominant factor controlling behaviour in the selection task is the highly abstract formal distinction between deontic and descriptive interpretation. But finding out how the details of the materials trigger the application of this distinction is a complex matter.

Psychologists need the abstractions provided by semantics as a basis for studying implementations in the mind. Logicians and linguists have much to gain from the data generated in the strange communications that go on in the psychological laboratory. These communications put subjects' interpretative skills under so much more stress than is customary, that they bring the interpretative issues to the surface.

In fact laboratory tasks have much in common with the curious communicative situation that is formal education and another benefit of the current approach is that it stands to reconnect the psychology of reasoning with educational investigations. With very few exceptions (e.g., Stanovich & West, 2000), psychologists of reasoning do not ask what educational significance their results have. They regard their theories as investigating 'the fundamental human reasoning mechanism' which is independent of education. On our account, the descriptive selection task is interesting precisely because it forces subjects to reason in vacuo and this process is closely related to extremely salient educational processes which are aimed exactly at equipping students with generalisable skills for reasoning in novel contexts more effectively. For example, the balance of required cooperative assumption of the background rule and adversarial test of the foreground rule in the descriptive selection task, is absolutely typical of the difficulties posed in the strange communications involved in examination questions. Many cross cultural

observations of reasoning can be understood in terms of the kinds of discourse different cultures invoke in various circumstances. The discourses established by formal education are a very distinctive characteristic of our culture (see e.g., Bloom & Broder, 1950; Hill & Parry, 1989).

There is often held to be something of a crisis in education in teaching these very reasoning and thinking skills (see Stenning, 2002 for an extended discussion). The prejudice against logically based accounts of human reasoning cuts off the insights of psychology from application to the educational problem. The community who teach reasoning are often as allergic to formal semantics as are psychologists of reasoning, largely because of past simplistic attempts to apply formal theories in monolithic ways. Now that logic is less monolithic, the fields cannot afford to continue avoiding each other.

Notes

1. There are a great variety of specific deontic stances which all share this feature that they deal in what is ideal relative to some criterion.
2. 'Language' should not be taken to be too literally here, since we do not want to exclude systems for reasoning with diagrams.
3. The following list with comments reflects the logician's practice. Textbooks are typically devoted to single, or at most a few, systems, and do not treat the matter in this generality. An exception is Gabbay's *Elementary Logics* (1998) although it has a more syntactic perspective than the one advocated here.
4. It is not possible here to describe these examples, which often have to do with 'plausible inference'. A worked-out example can be found in van der Does and van Lambalgen (2000). As a good analogy, the reader may think of statistical inference: on the basis of a statement ψ about a *sample S*, one concludes a statement φ about a *population P* from which *S* was drawn. Hence the set of models relevant for the premises (samples) is disjoint from the set of models relevant to the conclusion (populations).
5. In the following, $\triangleright$ is the 'makes true' or 'supports' relation introduced in Section 1.1.
6. Emphasis added.
7. In the psychological literature one may sometimes find a superficially similar distinction between descriptive and deontic conditionals. See, for example, Oaksford and Chater (1994), who conceive of a deontic conditional as material implication plus an added numerical utility function. The preceding proposal introduces a much more radical distinction in logical form.
8. Our student Marian Counihan ran an additional 10 subjects in October 2003. The results show a striking resemblance to the ones reported here. The interested reader may contact the authors for further information.
9. In that very small sample the baseline testing prior to tutoring showed no simple increase in proportion of completely correct performance (on the classical model), although tutoring in the two-rule task was more effective than in the classical task.
10. The actual experimental set up is much more complicated and not quite comparable to the experiments reported here.

11. Four combinations, because the deictic back/face reading of 'one side ... other side' appeared to be too implausible to be considered. But see below.

Acknowledgments

References

Bloom, B. S., & Broder, L. J. (1950). *Problem solving processes of college students.* Chicago: University of Chicago Press.

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

Cheng, P., & Holyoak, K. (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.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.

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*, 193–199.

Evans, J., Newstead, S., & Byrne, R. (1993). In S. Newstead & R. Byrne (Eds.), *Human reasoning: The psychology of deduction*. Hove: Lawrence Erlbaum.

Evans, J., & Over, D. (1996). *Rationality and reasoning*. Hove: Psychology Press.

Fiddick, L., Cosmides, L., & Tooby, J. (2000). The role of domain-specific representations and inferences in the Wason selection task. *Cognition*, *75*, 1–79.

Fillenbaum, S. (1978). How to do some things with it. In Cotton & Klatzky (Eds.), Semantic functions in cognition. Lawrence Erlbaum Associates.

Gabbay, D. (1993). A general theory of structured consequence relations. In P. Schroeder-Heister & K. Došen (Eds.), *Substructural logics*. Oxford: Clarendon Press.

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

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition*, *43*, 127–171.

Girotto, V., Kemmelmeier, M., Sperber, D., & van der Henst, J.-B. (2001). Inept reasoners of pragmatic virtuosos? Relevance in the deontic selection task. *Cognition*, *81*, B69–B76.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts*. London: Academic Press.

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

Goodman, N. (1954). *Fact, fiction and forecast*. London University Press.

Henle, M. (1962). On the relation between logic and thinking. *Psychological Review*, *69*, 366–378.

Hill, C., & Parry, K. (1989). Autonomous and pragmatic models of literacy: Reading assessment in adult education. *Linguistics and Education*, *1*, 233–283.

Hodges, W. (1993). The logical content of theories of deduction. Commentary on Johnson-Laird & Byrne Deduction. *Behavioural and Brain Sciences, 16*(2), 353–354.

Hughes, M. (1966). The use of negative information in concept attainment. Ph.D. thesis, University of London.

Johnson-Laird, P., & Byrne, R. (2002). Conditionals: A theory of meaning, pragmatics and inference. *Psychological Review.*

Johnson-Laird, P., Legrenzi, P., Girotto, V., & Legrenzi, M. (2000). Illusions in reasoning about consistency. *Science*, *288*, 531–532.

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

Johnson-Laird, P., & Savary, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition*, *71*(3), 191–229.

Kirby, K. (1994). Probabilities and utilities of fictional outcomes in Wason's selection task. *Cognition*, *51*(1), 1–28.

Manktelow, K., & Over, D. (1990). *Inference and understanding: A philosophical perspective*. London: Routledge.

Margolis, H. (1988). *Patterns, thinking, and cognition: A theory of judgement*. University of Chicago Press.

Newstead, S. (1995). Gricean implicatures and syllogistic reasoning. *Journal of Memory and Language*, *34*, 644–664.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608–631.

Platt, R., & Griggs, R. (1993). Facilitation in the abstract selection task: The effects of attentional and instructional factors. *Quarterly Journal of Experimental Psychology-A*, *46*(4), 591–613.

Rips, L. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, *57*, 31–95.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford: Blackwells.

Stanovich, K., & West. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioural and Brain Sciences, 23*, 645–726.

Stenning, K. (2002). *Seeing reason. Image and language in learning how to think.* Oxford: Oxford University Press.

Stenning, K., & Cox, R. (submitted). Rethinking deductive tasks: Relating interpretation and reasoning through individual differences.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: logic and implementation. *Cognitive Science*, *19*, 97–140.

Stenning, K., & van Lambalgen, M. (2001). Semantics and psychology: Wason's selection task as a case study. *Journal of Logic, Language and Information 10*(3), 273–317.

Stenning, K., & van Lambalgen, M. (submitted). *A working memory model of relations between interpretation and reasoning*, http://www.hcrc.ed.ac.uk/keith/InterpretationandReasoning/suppression.pdf.

Stenning, K., & Yule, P. (1997). Image and language in human reasoning: A syllogistic illustration. *Cognitive Psychology*, *34*, 109–159.

Stenning, K., Yule, P., & Cox, R. (1996). Quantifier interpretation and syllogistic reasoning. In G. W. Cattrell (Ed.), *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 678–683). Erlbaum: Maweh, N.J.

van der Does & van Lambalgen (2000). A logic of vision. *Linguistics and Philosophy 23*, 1–92.

van Duyne, P. (1974). Realism and linguistic complexity. *British Journal of Psychology*, *65*, 59–67.

Wason, P. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*, 273–281.

Wason, P. (1987). Problem solving. In R. Gregory (Ed.), *Oxford companion to the mind* (pp. 641–644).

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

Wason, P., & Johnson-Laird, P. (1969). Proving a disjunctive rule. *Quarterly Journal of Experimental Psychology*, *21*, 14–20.

Wetherick, N. E. (1970) On the representativeness of some experiments in cognition. *Bulletin of the British Psychological Society*, *23*, 213–214.