3 A Little Logic Goes a Long Way

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., Henle [122], Gebauer and Laming [92], Newstead [201], Byrne [28]), 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 into cognitive theory generally, and human reasoning in particular. This chapter exemplifies this general claim in the domain of Wason's selection task [295], 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, etc.) 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 chapter. 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, honorable exceptions which we will consider in our discussion.

In this chapter we take Wason's selection task and argue that the mental processes it evokes in subjects are, quite reasonably, dominated by interpretive 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 (Johnson-Laird and Byrne [144]), relevance theory (Sperber, Cara, and Girotto [252]), "evolutionary psychology" (Cosmides [47]), rational analysis (Oaksford and Chater [205]). 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 leading to 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.

3.1 The Mother of All Reasoning Tasks

Wason's original task was presented by means of a form as depicted in figure 3.1. The reader, who has probably seen the task before, should realise that this is all the information provided to the subjects, in order to appreciate the tremendous difficulty posed by this task. We will later present a variant of the task (the "two rule task") which may recreate in the cognoscenti the feelings of perplexity experienced by the untutored subject.

> 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 |

Figure 3.1 Wason's selection task

This experiment has been replicated many times. If one formulates the rule

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

as an implication $p \rightarrow q$, then the observed pattern of results is typically given as in table 3.1.

Thus, the modal response[1] (around half of the undergraduate subjects) is to turn A and 4. Very few subjects turn A and 7. Wason (and, until fairly recently, the great majority of researchers) assumed, without considering alternatives, that the correct performance is to turn the A and 7 cards only. Oaksford and Chater's inductive rational choice model [205] was the first to challenge this assumption, by rejecting deductive models entirely – more on this below. Wason adopted this criterion of good reasoning from a classical logical interpretation of the rule.

In a very real sense, however, Wason got his own task wrong in stipulating that there was a particular "obviously correct" answer. This holds on the assumption that the rule is interpreted using the material implication, but nothing in the experimental material itself forces this interpretation. Undergraduate subjects come to this task with their own interpretation of the conditional "if ... then," and more often than not this interpretation is as a defeasible rule robust to exceptions. In this case, 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.[2] 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. Furthermore, subjects are not accustomed to reflecting on their language use and also 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 antilogical. In our discussion we review some of the stances toward logic exhibited by the prominent cognitive theories that have made claims about the selection task, and appraise them from the viewpoint advocated here.

1. The mode of a data sample is the element that occurs most often in the sample.
2. It is true that Wason's instructions explicitly state that the rule applies only to the four cards, but, as we shall see below, this is not a reasonable restriction on one dominant interpretation, and is widely ignored.

Table 3.1 Typical proportions of choices in the selection task

p	p, q	$p, \neg q$	$p, q, \neg q$	misc.
35%	45%	5%	7%	8%

3.2 A Preliminary Logical Distinction

The program of empirical investigation that ensued from Wason's original experiment can be seen as a search for those contents of rules that 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.

For example, when the original "abstract" (i.e., descriptive) form of the selection task proved so counterintuitively hard, attention rapidly turned to finding materials that made the task easy. Johnson-Laird, Legrenzi, and Legrenzi [142] 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. They described the facilitation in terms of *familiarity* of the materials. Griggs and Cox [109] showed that reasoning about a drinking age law ("if you drink alcohol here, you have to be over 18") was easy. Wason and Green[296] similarly showed that a rule embedded in a "production-line inspection" scenario (e.g., "If the wool is blue, it must be 4 feet long") also produced good performance. Cosmides [47, 48] went on to illustrate a range of 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. Note that nonsocial contract examples such as the "wool" example just quoted were prominent in the literature before Cosmides made these claims. 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, and Tooby [81]).[3] These observations of good reasoning were 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 and Tooby [48,p. 166])

3. Cummins [53] has argued, against this proliferation, that the innate module concerned is more general and encompasses all of deontic reasoning.

Johnson-Laird equally claims that the effect of content on reasoning contradicts the formal-logical model of reasoning:

> [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 [146,p. 225])

Wason himself, discussing an example from [299] which used the rule: "If I go to Manchester I go by train," 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*... ([300,p. 643]; emphasis added)

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 *logical form*, and we explain the effect of this interpretive choice in terms of the many problems descriptive interpretation creates in this task setting, as contrasted with the ease of reasoning 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, as described in the previous chapter, state how matters *should* be according to some (perhaps legal) law or regulation, or preference.[4] 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.[5]

Returning to the examples given, we see that a rule like "if you drink alcohol here, you have to be over 18" is deontic, not descriptive. If everybody drinking whiskey in a particular pub is under 18, the rule may be massively violated, but it is still in force. We have a different situation if the rule is intended to be descriptive, an intention which can be suggested, but not fixed, by the formulation "if someone drinks alcohol here, he is over 18." In this case a primary school boy sipping his whiskey may be taken to falsify the rule. The other examples can be analysed similarly.

4. There are a great variety of specific deontic stances which all share the feature that they deal in what is ideal relative to some criterion.

5. This is an important invocation to be born in mind by the formalist concerned to contribute to experimental investigation: "Use the bluntest formal analysis which is sharp enough to make the necessary distinctions!"

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 descriptive 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 generalization or a legal prescription. Conversely, subjunctive verbs and modals are often interpreted descriptively. (e.g., in the sentence "If it is 10 am, 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 auxiliary verbs in rules. A combination of rule, content, and subjects' knowledge influences whether to assign a deontic or descriptive *form*.

Our proposal about the selection task at its simplest is that the semantic difference between descriptive and deontic conditionals leads to a processing difference when these conditionals are used in the selection task. In barest outline, the semantic difference is this. Deontic conditionals such as the drinking age law cannot be false, they can only be violated. Hence turning a card never affects the status of the rule. Now, whether a given customer violates the drinking age law is independent of whether another customer does so. This seems a trivial observation, but a comparison with the case of descriptive rules will show its importance.

A descriptive rule can be true or false, and here the purpose of turning cards is to determine which is the case; in this case, unlike the previous one, the result of turning a card can affect the status of the rule. A consequence of this is a form of dependence between card choices, in the following sense: turning A to find 7 already decides the status of the rule by showing it to be false; it does not seem necessary to turn the 7 card as well. On the other hand, if upon turning A one finds a 4, it is necessary to turn the 7 as well. We will see that this situation is confusing to subjects, who sometimes believe that the task is unsolvable because of this dependence. It is not, of course, once one realises that the cards are not real but depicted, and hence cannot be turned at all. These blunt semantic differences mean that the original descriptive (abstract) task poses many difficulties to naive reasoners not posed by the deontic task.[6] A formal analysis of the differences will be given below, but these indications should suffice for the present.

We will derive a variety of particular difficulties to be expected from the inter-

6. Previous work has pointed to the differences between the deontic and descriptive tasks (e.g., Manktelow and Over [180]; Oaksford and Chater [205]). Cheng and Holyoak [37] developed the theory that success on deontic selection tasks was based on *pragmatic reasoning schemata*. 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 analyzing abstractly the contrasting difficulties which descriptive conditionals pose in this task.

action 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 generalization about reasoning in this task, and an explanation of why that generalization holds. It also strongly supports the view that subjects' problems are highly variable and in doing 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 interpretive 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 or 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 interpretive 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. The take-home message is: semantics supplies an essential theoretical base for understanding the psychology of reasoning.

The plan of the remainder of the chapter 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 interpretive perspective for cognitive theories more generally.

3.3 Logical Forms in Reasoning

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 the modern conception of logic outlined in chapter 2. By this, we do not just mean that modern logic has come up with other competence models besides 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 behavior 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 what is not, 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 parameters 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 *to* 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.

We have seen in chapter 2 that the parameters characterizing logical form come in at least three flavors: pertaining to syntax, to semantics, and to the notion of logical consequence. In short, therefore, the interpretive 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 postmodern hermeneutics. This doomful view may be partly due to earlier psychological invocations of interpretational defenses against accusations of irrationality in reasoning, perhaps the most cited being Henle [122]: "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 was, 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,[7] 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 lamppost. 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.

7. Though see [74] to see how negation is not far away.

3.4 Logical Forms in the Selection Task

We now apply the preceding considerations to the process of assigning a logical form to the standard selection task. An important component of this process is determining a meaning of the conditional, but this is not all there is to logical form in the selection task. To see this, it is useful to reformulate the task as an *information processing task* as intended by Marr [183].[8] Here we run into an immediate ambiguity: the information–processing task may be meant as the one intended by the experimenter, or as understood by the subject. In both cases the input consists of the form with instructions, but the output can be very different.

The intentions of the experimenter are clear: the output should consist of checks under the cards selected. The output in the subject's interpretation of the task can vary considerably. Some subjects think the output can be a plan for showing the rule to be true or false. Other subjects interpolate a process of information gathering and view the task as "what information do I require to decide the rule, and how do I obtain that information." This gets them in considerable trouble, since they tend to rephrase the minimality condition "do not turn unnecessary cards" as the condition "determine the minimum information necessary to decide the rule," which is importantly different. In this section we will describe several formal models corresponding to different understandings of the task.

Wason had in mind the interpretation of the conditional as a truth–functional implication, which together with bivalence yields the material implication. Truth–functional, because the four cards must decide the truth–value of the conditional; bivalent, because the task is to determine *truth or falsity* of the conditional, implying that there is no other option. 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 her own most prominent meanings for the key terms involved, the subject may deduce that the conditional is to be interpreted truth–functionally, 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.

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 the "abstract" rule 1 actually needs predicate logic,

8. Marr's ideas will be explained in greater detail in chapter 5 in section 5.1.4 and in chapter 11.

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:
$V(x, y)$ "x is on the visible side of card y"
$I(x, y)$ "x is on the invisible side of card y"
$O(x)$ "x is a vowel"
$E(x)$ "x is an even number"
and the rule is then translated as the following pair:

$$\forall c(\exists x(V(x,c) \land O(x)) \to \exists y(I(y,c) \land E(y)))$$

$$\forall c(\exists x(I(x,c) \land O(x)) \to \exists y(V(y,c) \land E(y)))$$

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

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

or even a reversed conditional

$$\forall c(\exists x(V(x,c) \land E(x)) \to \exists y(I(y,c) \land O(y))).$$

This very interesting phenomenon will be studied further in section 3.5.5.[9]

For simplicity's sake, in the following we will focus only on subjects' problems at the level of propositional logic.

3.4.1 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, in contrast to its familiarity. 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.*

Descriptive Interpretation of the Task

We assume first that the subject views the task descriptively, as determining whether the rule is true or false of the four cards given, no exceptions allowed.[10]

9. The fact that Gebauer and Lamming [92] accept this reversed reading in this context provides a timely reminder that subjects and experimenters are sometimes equally prone to accept interpretations which would be held to be quite incongruous in a more neutral context.
10. Some subjects have a descriptive interpretation which does allow exceptions; see section 3.5.2.

We also assume that cards with false antecedent (i.e., a consonant) are viewed as complying with the rule[11] This dictates an interpretation of the descriptive conditional as material implication.

Wason viewed his task in terms of bivalent classical logic, but even in the descriptive case this holds only for an omniscient being, who can view both sides of the cards. The human subject who comes to this task is at first confronted with a lack of information, and may apply a logic which is appropriate to this situation: a semantics which is sensitive to the information available to the subject. It will turn out that there are two subtly different ways of taking into account the incomplete information, which will be seen to correspond to different behaviors of subjects. In the process of formalization, we will also have occasion to introduce two different notions of truth that persistently confuse subjects.

The first formalization of the selection task views it as an information–processing task whose output is the information the subject requires for deciding the rule. Suppose for simplicity that the letters on the cards can only be A, K, and the numbers only 4,7. Define a model, consisting of "information states" as follows. For brevity, we will sometimes call information states "states." There are four states corresponding to the visible sides of the cards; denote these by $A, K, 4, 7$. These correspond to incomplete information states. Then there are eight states corresponding to the possibilities for what is on the invisible side; denote these by $(A,4)$, $(A,7)$, $(K,4)$, $(K,7)$, $(4,A)$, $(4,K)$, $(7,A)$, and $(7,K)$. These correspond to complete information states (about a single card). This gives as domain W of the model twelve states in all, each pertaining to a single card. Starting from W one may define the set W_4 consisting of all consistent information states relating to the four cards simultaneously. W_4 contains sets such as $\{(A,4), K, 4, 7\}$, $\{A, K, (4,A), (7,A)\}$, $\{(A,4), K, (4,A), 7\}$, or $\{(A,7), (K,4), (4,K), (7,K)\}$. On W_4 one can define an ordering $\leqslant$ by $w \leqslant v$ if the information contained about a given card in v is an extension of, or equal to, the information about that card in w. So we have, e.g., $\{(A,4), K, 4, 7\} \leqslant \{(A,4), K, (4,A), 7\}$ and $\{A, K, 4, 7\} \leqslant \{(A,7), (K,4), (4,K), (7,K)\}$, but $\{(A,4), (4,A), K, (7,A)\} \nleqslant \{A, (4,A), K, (7,A)\}$.

To represent the different relations of truth that cards and rule can bear toward each other, we introduce a support-relation $\Vdash$ and the standard "makes true" relation $\models$. The latter relation is in a sense symmetric: if for a model $\mathcal{A}$ and a formula φ one has $\mathcal{A} \models \varphi$ one may say equivalently that $\mathcal{A}$ makes φ true, or that φ is true of $\mathcal{A}$. This is different for the support-relation $\Vdash$, holding between a "piece of information" v and a formula φ: $v \Vdash \varphi$ must be read as the asymmetric relation "v contains evidence for φ." It is the interplay between the asymmetric and the symmetric relation that causes many subjects a headache.

11. It would not be appropriate to give the implication the truth–value "undecided" (as in Kleene's three–valued logic) in this case, since no amount of additional information can affect this truth–value.

The support-relation $w \Vdash \psi$ is defined between states in W_4 and formulas ψ as follows. Let p be the proposition "the card has a vowel," and q the proposition "the card has an even number." The referent of the expression "the card" is determined by the information state on which p, q are interpreted; we assume that in an expression $p \wedge (\neg)q$ the referent of "the card" in q is bound by that in p. Then we have

1. $A \Vdash p$, $K \Vdash \neg p$, p undecided on 4 and 7 (i.e., neither $4 \Vdash p$ nor $4 \Vdash \neg p$, and similarly for 7);
2. $4 \Vdash q$, $7 \Vdash \neg q$, q undecided on A and K;
3. $(A, 4) \Vdash p \wedge q$, $(A, 7) \Vdash p \wedge \neg q, \ldots (4, A) \Vdash p \wedge q$, $(4, K) \Vdash \neg p \wedge q$, etc.

For a state v in W_4, define $v \Vdash p \wedge \neg q$ as: "there is a card (x, y) in u such that $(x, y) \Vdash p \wedge \neg q$." We say that the rule is supported by a piece of information v, and write $v \Vdash p \rightarrow q$, if $v \nVdash p \wedge \neg q$. Lastly we say that v makes the rule true, and write $v \models p \rightarrow q$, if for all $u \geqslant v : u \Vdash p \rightarrow q$. Clearly $v \models p \rightarrow q$ implies $v \Vdash p \rightarrow q$, but the converse does not hold, and this is one source of confusion for subjects, as we will see in section 3.5.2. In this "information-seeking" version of the task the subject now must compute the information states that decide the rule. A combinatorial exercise involving the truth–table for $\rightarrow$ shows that these are: $\{(A, 7), K, 4, 7\}$, $\{A, K, 4, (7, A)\}$, $\{(A, 4), K, 4, (7, K)\}$ and their extensions. Suppose the subject now views the task as: "what information must *I* gather in order to decide the rule," and interprets the instruction not to turn unnecessary cards in this light, thus looking for minimal information states only. The subject must then perform an action, or actions, which bring her from $\{A, K, 4, 7\}$ to one of the desired minimal information states. The trouble is that sometimes turning a single card suffices to achieve a minimal information state, and that sometimes turning two cards is necessary, and it depends on the unknown hidden side of the cards which situation one is in. Subjects interpreting the task in this way therefore think that it is unsolvable. To make the task solvable, a different interpretation is needed, in which the subject does not think in terms of the information which must be *gathered*, but in terms of information which becomes *available*. The formalization appropriate to this interpretation involves a game in which the subject plays against an adversary who makes the information available; the subject's optimal choice corresponds to a winning strategy in this game. A formal characterization goes as follows.

The game is played between two players, I (the subject) and II (the experimenter). The game is played on sixteen boards simultaneously, corresponding to what can be on the back of the cards whose visible sides are $A, K, 4, 7$. Player I selects a set of cards which she claims will decide the rule on every board. Player II has two options:

(a) II picks a card outside I's selection and chooses a board;
(b) II picks a card in I's selection and chooses two boards.

In case (a), II's move is winning if I's selection makes the rule true, but II's card makes it false on the chosen board. In case (b), II's move is winning if II's selected card has different values on the two boards, but the rule is true on both. Player I wins if II does not win. Clearly a winning strategy for player I consists in the selection of A and 7. Less, and player II can win via a move of type (a); more, and II can win via a move of type (b). To compute the winning strategy is again a combinatorial exercise.

It is essential for this version of the task that the game is played on sixteen boards simultaneously. If the game were to be played on a single board about which player I knows nothing, the concept of strategy would not make sense, and we find subjects saying that it is only luck that a selection turns out be relevant; see for instance, subject 5 in section 3.5.3. Hence if the game corresponding to the selection task is formalized as being played on one board only, the subject must consider the task to be unsolvable.

There is thus much more to the descriptive selection task than a simple computation in propositional logic which subjects cannot do because they are partly irrational, or have no general logical competence. Indeed, that subjects find the purely propositional part easy is witnessed by the fact that they do not have trouble evaluating the impact of a card correctly (e.g., subject 3 in the introduction to section 3.5). It is the choosing that creates the difficulty, and here the instructions provide little guidance on how to construct the proper information–processing task, or they may even actively interfere with the construction.

Deontic Interpretation of the Task

The logic appropriate to a deontic interpretation of the task is very different. One difference is in the structure of the models associated with deontic statements.

As we have seen in chapter 2, a deontic model $\mathcal{A}$ is given by a set of "worlds" or "cases" W, together with a relation $R(w, v)$ on W intuitively meaning: "v is an ideal counterpart to w." That is, if $R(w, v)$, then the norms posited in w are never violated in v. The relation R is used to interpret the modal operator O ("it ought to be the case that"). On such a model, we may define a deontic conditional by writing $O(p \to q)$, but it is convenient to introduce a special notation for this conditional, namely $p \prec q$, and to define for any world $w \in W$

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

and

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

The satisfaction relation for atomic propositions is the same as that in the descriptive case.

The definition thus introduces an additional parameter R. If W is the set of worlds defined above, 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 $R(4,(4,K))$.[12] R encodes the evaluation of each card against the norm, and this is what subjects can easily do.

As the reader will have no trouble verifying, the deontic model $(W, R, \models)$ then satisfies $p \prec q$, i.e., for all w in W: $w \models p \prec q$. For example, to verify that $A \models p \prec q$, one notes that the only world v satisfying $R(A, v)$ is $(A, 4)$, which satisfies both p and q. That is, in contrast to the previous case the rule is true from the start, hence there is no need to gather evidence for or against the rule, and the conflicts between $\models$ and $\Vdash$, and between two views of information, cannot arise in this case. The information-processing task becomes rather different: the output is the set of cases which possibly violate the norm, and inspection of the definition of R shows that only A and 7 are candidates. Turning 7 to find A just means that $(7, A)$ is not an ideal counterpart to 7, in the sense that it does not satisfy the norms holding in 7. The computation is accordingly just a simple lookup. The set W_4 and the strategic choices it gives rise to do not enter into the picture.[13]

Deontic Connectives

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 one side of the cards *and* there is an even number on the other side,[14]

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 4.1.5). One reason for this behavior is given by subject 22 in section 3.5.6, 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,

$$w \models p \sqcap q \text{ iff for all } v \text{ such that } R(v,w)\text{: } v \models p \wedge q.$$

12. We shall assume that if v is an ideal counterpart to w, then v is maximally ideal, that is, v is the ideal counterpart of itself. Thus we assume $\forall w \forall v(R(w,v) \rightarrow R(v,v))$.

13. In the psychological literature one may sometimes find a superficially similar distinction between descriptive and deontic conditionals. See, e.g., Oaksford and Chater [205], 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.

14. Emphasis added.

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 . and 4, to see whether perhaps *these* worlds are ideal.

Domains

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 a 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 3.5.2 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.

Other Logical Forms

Some subjects believe a conditional allows exceptions, and cannot be falsified by a single counterexample (see section 3.5.2). 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, but not all, 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. These robust default conditionals were mentioned above in chapter 2, section 2.3 and will be used below in chapter 7 to model discourse interpretation in the suppression task.

This concludes our survey of what is involved in assigning logical form in this task. 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.

3.5 Giving Subjects a Voice

A standard selection task experiment consists in giving subjects a form which contains the instructions and shows four cards; the subjects then have to mark the cards they want to select. The type of data obtainable in this way is highly abstracted from the reasoning process. The subjects' approach to the task may be superficial in the sense of not engaging any reasoning or comprehension process which would be engaged in plausible real-world communication with the relevant conditionals. One loses information about subjects' vacillations (which can be very marked) and thus one has little idea at what moment of their deliberations subjects make a choice. It is also possible that the same answer may be given for very different reasons. Furthermore, the design implies that the number of acceptable answers is restricted; for instance, some subjects are inclined to give an answer such as "A or 4," or "any card," or "can't say, because it depends on the outcomes," and clearly the standard design leaves no room for such answers. Early on, Wason and Johnson-Laird [297] investigated the relationship between insight and reasoning by also using dialogue protocols. They distinguished two kinds of feedback: (1) feedback from hypothetical turnings –"suppose there is an A on the back of the 7, what would you then conclude about the rule?"; (2) actual feedback in which the subject turns the 7 card and finds the A – "are you happy that you did/didn't select the 7 card?." It seems to us that this type of design is much more conducive to obtaining information about the whys and wherefores of subjects' answers.

We report here excerpts from Socratic tutorial dialogues with subjects engaged in the task, to illustrate the kinds of problems subjects experience.[15] Observational studies of externalized reasoning can provide prima facie evidence that these problems actually are real problems for subjects, although there is, of course, the possibility that externalizing changes the task. Only a controlled experiment can provide evidence that the predicted mental processes actually do take place when subjects reason in the original noninteractive task. A controlled experiment whose hypotheses derive from a combination of the present material with a logical analysis will be reported in the next chapter.

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. But rather than turn these observations into a quantitative study of the dialogues which would still only bear on this externalized 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

15. Some of these excerpts were reported in [264] and [265[. Others come from a tutorial experiment performed by the our student Marian Counihan.

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.

As an appetiser we present two examples of dialogues, both of kinds observed by Wason. The first example would be considered a case of "irrationality" by Wason (but not by us), the second shows a possibly related and initially perplexing dissociation between logical evaluation and selection of cards which Wason also found striking.

The first example shows that subjects may fail to understand the implication of the 7/A combination. Here, as in the sequel, we denote by "7/A" the card which has 7 on the visible face and A on the invisible back.

Subject 14
S. I would just be interested in As and 4s, couldn't be more than that.
E. So now let's turn the cards, starting from right to left. [Subject turns 7 to find A.]
E. Your comments?
S. It could be an A, but it could be something else. *E.* So what does this tell you about the rule?
S. About the rule ... that if there is an A then maybe there is a 7 on the other side.
E. So there was a 7.
S. But it doesn't affect the rule.

The second example shows that a subject sometimes hypothesises (or discovers) an A on the back of the 7, and notes that this would mean the rule was false of the card, but then declines to choose the card (or revise an earlier failure to choose it).

Subject 3
E. OK. Lastly the 7.
S. Well I wouldn't pick it.
E. But what would it mean if you did?
S. Well, if there is an A then that would make the rule false, and if there was a K, it wouldn't make any difference to the rule.

Following the theory outlined in chapter 2 and section 3.3, 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 interpretational 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 vs. "truth" for a case
Cards as viewed as a sample from a larger domain
Obtaining evidence for the rule vs. evaluation of the cards
Existential import of the conditional
Subjects' understanding of propositional connectives generally

Interestingly, most of these problems simply cannot occur on a deontic interpretation of the task, and we take this to be the reason why performance on this task is so much 'better' than on the descriptive task.

3.5.1 The Design of the Tutorial Experiment: High-Energy Phenomenology[16]

The experiment to be reported in the next section[17] consisted of two parts. First we gave subjects a booklet with the standard Wason task and the two–rule task (for which see below), which also contained a so–called *paraphrase* task, in which the subjects were asked to judge entailment relations between sentences involving propositional connectives and quantifiers. This task continues the classical work of Fillenbaum [82] 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,[18] but we will see below that these judgments can be logically startling. The results of this task gave us some information about subjects' understanding of logical connectives, which could then be related to their performance in the selection task.

The second part of the experiment consisted of a series of dialogues with the subject while she or he was engaged in solving the Wason task or one of its variants. The dialogues were recorded on video and transcribed. In the case of the standard task, the setup was as follows. The subject received a form giving the same instructions as we saw in figure 3.1, except that the pictures of the card were replaced by real cards, again showing A,K, 4, and 7. We first asked the subjects to select cards. We then asked them to reflect on what might be on the other side, given the instructions, and to evaluate the imagined outcomes with respect to the truth–value of the rule. Subjects were then allowed to revise

16. We first came across this phrase in an ad for a physics job, but it seemed to capture the feeling of Wason's materials scattering into a thousand interpretations when they impinged on the subjects' expectations, as well as the amount of effort expended by both E and S in struggling to understand events.
17. The experiment was performed in 1999 in Edinburgh by us together with Magda Osman.
18. Although not quite, as it brings to the fore issues about the existential import of the conditional, to which we will return in section 3.5.6.

their initial selection. Lastly, we asked them to turn each card and to explain to us what the result implies for the truth–value of the rule. We also followed this procedure for variants of the standard task, such as the one explained next.

A Two–Rule Task

This task, whose standard form is given as figure 3.2, is the first in a series of manipulations which try to alleviate some of the difficulties subjects have in interpreting the task. The classical logical competence model specifies that correct performance is to turn just the 3 card. We conjectured that 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. Below we give excerpts from the tutorial dialogues which highlight these difficulties. 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. In this case, both U and I carried an 8, 8 carried an I, and 3 a U.

> 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 an 8 or a 3 on one of its sides and a U or I on the other.
> Also below there are two rules which apply only to the four cards. It is given that exactly one rule is true. Your task is to decide which if any of these four cards you *must* turn in order to decide which rule is true. Don't turn unnecessary cards. Tick the cards you want to turn.
>
> 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.*
>
> **Cards:**
>
> | | | | |
> |---|---|---|---|
> | U | I | 8 | 3 |

Figure 3.2 Two–rule task.

The task is pragmatically somewhat peculiar in that the two rules different in the antecedent, not the consequent, and are still said to be mutually exclusive. Naturally occurring mutually exclusive rules seem to have the same antecedent but different consequents. The antecedent of a conditional often acts as a topic (in the linguistic sense), and the two conditionals then say something different of this topic. It is much less common to have two topics, each corresponding to an antecedent. Occasionally one therefore observes pragmatic normalization in the dialogues, which inverts the conditionals to "if 8 then U" and "if 8 then I."

We are now in a position to present subjects' musings, insights and perplexities while working through the various tasks.

3.5.2 Subjects' Understanding of Truth and Falsity

'Truth' and 'falsity' are among the most important parameters to be set, and the reader should recall from chapter 2 that they can be set independently: only classical logic forces "not true" to be the same as "false." This stipulation corresponds to a definition of semantics, in which one defines only "true on a model" ($\models$), not "false on model" ($\not\models$). It is, however, equally possible to give a recursive definition of semantics in which $\models$ and $\not\models$ are defined by simultaneous recursion. We therefore give 'true' and 'false' separate headings.

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. Semantically this means, for a given model $\mathcal{A}$ that if not $\mathcal{A} \not\models p \rightarrow q$, then $\mathcal{A} \models p \rightarrow q$. Interestingly, some subjects refuse to be moved by the explicit instruction, insisting that "not-false" is not the same as "true." These subjects are thus guided by some nonclassical logic.

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. [First] 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 1 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" (this answer may reflect the pragmatic normalization of the two conditionals referred to above). 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 1?
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 law-like character to the conditional, then the absence of counterexamples does not suffice to establish truth; a further causal connection between antecedent and consequent is necessary.[19] Since the truth of $p \rightarrow q$ cannot be established on the model w in W_4, the semantics implicitly adopted by these subjects for $p \rightarrow q$ is a form of intensional semantics, following the terminology introduced in chapter 2. That is, to determine $w \models p \rightarrow q$ one has to refer to information extraneous to w, for example a much larger population of cards. A subject faced with this difficulty will be unable to solve the task as given, because of lacking information.

Let us note here that this difficulty is absent in the case of deontic rules such as

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

19. Such a causal connection is present in the "production line scenarios" which ask the subject to imagine that there is a machine producing cards which on one side ..., etc.

Such a rule cannot, and in fact need not, be shown to be true by examining cases; its truth is given, and the subjects need only establish that it is not violated. So in the deontic case, subjects have no worries about the meaning of "truth."

Another twist to the concept of truth was given by those subjects, who, 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. Again this leads to an intensional semantics for $p \rightarrow q$.

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; it is possible to give a separate definition of $w \not\models p \rightarrow q$.

The paraphrase task alluded to above showed that a conditional $p \rightarrow q$ being false is often ($> 50\%$) interpreted as $p \rightarrow \neg q$! (We will refer to this property as *strong falsity*.) This observation is not ours alone: Fillenbaum [82] 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 $p \rightarrow \neg q$; for contingent universals (conditionals, such as the rule in the selection task, where there is no salient connection between antecedent and consequent) 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.

> *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 (i.e., that the logic is bivalent), this could reflect a conviction that the cards have been laid out according to *some* rule, instead of randomly. It is interesting to see what this interpretation means for card choices in the selection tasks. If a subject has strong falsity and applies the (classical) tautology $(p \rightarrow q) \vee (p \rightarrow \neg q)$, then (in the standard Wason task) *either* of the cards $A, 4$ can show that $p \rightarrow q$ is not-false, hence true. Unfortunately, in the standard setup '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:

Subject 10
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. 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.

Truth and Satisfaction

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 [205]. We will return to this interpretation in section 3.5.2 below. The other extreme is that each card defines a domain of its own: each card is to be evaluated against the rule independently.

This interpretation is the one suited to deontic conditionals, though it is also possible with descriptive interpretations and is sometimes encouraged by "seek violations" instructions (see [308]). What is perhaps most tantalizing reading the early literature on the task is how little the experimenters themselves noticed that deontic interpretation was critical. There was a good deal of attention to the effects of "seek violators" instructions, but these were interpreted against the background of Wason's focus on falsification, not on their effect of making the cards independent of each other, or of removing the ambiguity between judging the cases and judging the truth of the rule.

An intermediate position is that there are cards which by themselves suffice to determine the truth–value of the rule; we saw an instance of this while discussing examples of "strong falsity," where the A and 4 cards are each decisive. The phenomenon may be more general, however, a failure to appreciate the relativity of the relation "the rule is true of the card." That is, even if a card satisfies the rule (what we called 'support' in section 3.4.1), it need not make it true.

Subject 10.
E. If you found an 8 on this card [I], what would it say?
S. It would say that rule 2 is true, and if the two cannot be true then rule 1 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 2 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 (equating truth with satisfaction) 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 1 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 2 would be true and rule 1 would be disproved. 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 2 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 would be to do a tutorial experiment where the A has a 4 on the back, and the 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.

It is clear that for a deontic rule no such confusion can arise, because the truth–value of the rule is not an issue.

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 is compatible with exceptions, to exceptionless truth.

In terms of logical form, this is the issue of the formal expression into which the natural language conditional is translated. As we observed in section 3.4, the proper formal correlate is a formula $p \wedge \neg e \rightarrow q$, where e is a proposition letter denoting an exceptional state of affairs, which can be given further content by a clause of the form "$\ldots \rightarrow e$."

Subject 18.
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 our data, the term "(possible) exception" was reserved for the $\neg q$ card; the p card qualified as a potential falsifier. We have no explanation for this phenomenon, but if it is pervasive, it would give yet another reason why subjects don't bother to look at the $\neg q$ card, even when they are clear about its logical meaning.

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 only to the four cards shown. 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 generalizations. So, to develop Goodman's example [102], "All the coins in my pocket this morning are copper" is a natural way to phrase a local generalization 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. Formally, this means that the set W of possible worlds introduced in section 3.4.1 must be replaced by a much richer structure.

A much more natural interpretation is that the four cards are a sample from a larger population. Indeed this is the point of purchase of Oaksford and Chater's proposals [205] that performance is driven by subjects' assumptions about the larger domain of interpretation. Some subjects raise this issue explicitly.

Subject 3.[20] [in standard Wason task; has chosen A, 4.]
S. Well in that case you would have to turn all the cards, if you couldn't work with just a start point. Because then if you turned.... take a random set of cards, imagine a random set of cards. If you had three A faces, and five 7 faces, but you couldn't have any assumption that that was a starting point, you would have to turn all the cards, because then you might get a 60%, 40% divide, and you would have to take an average, and say the rule isn't right, but the majority of cards suggest that if there's an A on one side then there's a 4. A likelihood, in that case.

20. In Marian Counihan's experiment.

E. But if there was a likelihood, what would that mean?
S. It wouldn't be a rule, it would be invalidated.

Here is another subject who thinks 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 2. [...] Well, there are two cards that verify rule 2, and only one card so far that verifies rule 1. Because if this [3] were verifying rule 1, it should be an I on the other side.
E. Let's turn [the 8].
S. OK, so that says that rule 2 is true as well, three of the cards verify rule 2 and only one verifies rule 1.
E. So you decide by majority.
S. Yes, the majority suggests rule 2.

It is interesting that 3/U is described as *verifying* rule 2, rather than *falsifying* rule 1; U→8 is never ruled out:

S. It's not completely false, because there is one card that verifies rule 1.

Asked to describe her thought processes, the subject later comments

S. Well, when there's two rules then you can't say that they should both be true because they are mutually exclusive ... so depending on which way the cards are there is basically a 50 per cent probability that either one is going to be true. ... With one rule I think it will be true or if it wouldn't be true, then it seems more likely that it would be true.

3.5.3 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.

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 2 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 [296] and discussed extensively by Margolis [182]. In its barest outline[21] 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, howevers, the frequency of the $\neg q$ response increases dramatically. From our point of view, this result is 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.

Getting Evidence for the Rule vs. Evaluation of the Cards

A related planning problem, which can, however, occur only on a nonstandard logical understanding of the problem, is the following. Some subjects interpret the instruction not to choose unnecessary cards as the injunction not to choose a card whose turning may yield a nondecisive outcome.

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.

21. The actual experimental setup is much more complicated and not quite comparable to the experiments reported here.

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 tradeoff between the 'information value' of a card and the penalty incurred by choosing it. Here is a subject very explicit about the trade-off.

Subject 3. [Standard Wason task; he realises the meaning of the 7 card, but that doesn't change his choice of only A]
S.... if there's an A on this side (pointing to the underside of the 7), it would invalidate the rule.
E. OK. So would that mean that you should turn the 7, or not?
S. Well you could turn the 7, but it says don't turn any cards you don't have to, and you only have to turn the A.
E. OK. So the 7 could have an A on it, which would invalidate the rule, but..
S. [interrupting] It could have, but it could also have a K on it, so if you turned that [the 7]) and it had a K, it would make no difference to the rule, and you would have turned a card that was unnecessary, which it says not to do.
E. But what if it had an A on it?
S. But what if it had a K on it?

Of course this does not yet explain the observed 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.

3.5.4 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 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 the possible sociopsychological 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 interpret 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 any 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.

3.5.5 Interaction between Interpretation and Reasoning

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 nonintended 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. That is, both "one side" and "other side" can be identified by direct pointing, whereas in the case of an anaphoric relationship the referent of "other side" depends on that of "one side." This possibility was investigated by Gebauer and Laming [92], 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.[22]

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 spelled 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 [264]. 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 to 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 the anaphora and the 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), (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?

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

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].

⋮
[Now in the standard Wason task]
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), (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.
E. 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 anaphor has 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 the back/face anaphora.

Subject 16. [experiments (1), (2), (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]

And here is the most striking example, in which the interaction can clearly be seen to go both ways.

Subject 23. [Standard Wason task]
S. Then for this card [4/K] the statement is not true.
E. Could you give a reason why it is not?
S. Well, I guess this also assumes that the statement is reversible, and if it becomes the reverse, then instead of saying if there is an A on one side, there is a 4 on the other side, it's like saying if there was a 4 on one side, then there is an A on the other.

⋮

E. Now we'll discuss the issue of symmetry, you said you took this to be symmetrical.
S. Well, actually it's effectively symmetrical because you've got this either exposed or hidden clause, for each part of the statement. So it's basically symmetrical.
E. But there are two levels of symmetry involved here. One level is the symmetry between visible face and invisible back, and the other aspect of symmetry is involved with the direction of the statement "if ... then."
S. Right, OK. so I guess in terms of the "if ... then" it is not symmetrical ... In that case you do not need that one [4], you just need [A]. ... [while attempting the two-rule task he makes some notes which indicate that he is still aware of the symmetry of the cards]
S. For U, if there is an 8 on the other side, then rule 1 is true, and you'd assume that rule 2 is false. And with I, if you have an 8, then rule 1 is false and rule 2 is true. ... [the subject has turned the U and I cards, which both carry 8 on the back, and proceeds to turn the 3 and 8 cards]
S. Now the 3, it's a U and it's irrelevant because there is no reverse of the rules. And the 8, it's an I and again it's irrelevant because there is no reverse of the rules. ... Well, my conclusion is that the framework is wrong. I suppose rules one and two really hold for the cards.
E. We are definitely convinced only one rule is true ...
S. Well ... say you again apply the rules, yes you could apply the rules again in a second stab for these cards [3 and 8] here.
E. What do you mean by "in a second stab?"
S. Well I was kind of assuming before you could only look at the cards once based on what side was currently shown to you. ... This one here [8[in the previous stab was irrelevant, because it would be equivalent to the reverse side when applied to this rule, I guess now we can actually turn it over and find the 8 leads to I, and you can go to this card again [3], now we turn it over and we apply this rule again and the U does not lead to an 8 here. So if you can repeat turns rule 2 is true for all the cards.
E. You first thought this card [3] irrelevant.
S. Well it's irrelevant if you can give only one turn of the card.

What's interesting in this exchange is that in the first experiment the variable, "symmetric" reading of the anaphora seems to trigger a symmetric reading of the implication, whereas in the second experiment asymmetric readings of the anaphora and the implications are conjoined, even though he was at first aware that the intended reading of the anaphora is symmetric. (The fact that the subject wants to turn the cards twice is evidence for the constant (asymmetric) reading of the anaphora.)

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, at least when asked to justify themselves. 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 were to decompose the anaphoric expression "one side... other side" into two deictic expressions "face/back" and "back/face" and were then to 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 [296] albeit in slightly different terms: their focus is on the distinction between a *unified* and a *disjoint* representation of the stimulus (i.e., a card). 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 to 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 [264]; here we discuss only their experiment 2.

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

(4) Whenever they are triangles, they are on black cards.

(5) 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 (5) was worse than for sentence (4) (for details, see Wason and Green [296,pp. 604-607]).

We would describe the situation slightly differently, in terms of the contrast between deixis and anaphora. Indeed, the experimental setup 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.

3.5.6 Subjects' Understanding of Propositional Connectives

We have discussed some aspects of the interpretation of the conditional above, in particular those which are connected with the interpretation of the task, such as the distinction between descriptive and deontic conditionals. This section is devoted to some of the wilder shores of the semantics of conditionals: the existential import of the conditional, and the related interpretation of the conditional as a conjunction.

Existential Import of the Conditional

In a second set of tutorial experiments[23] all the interpretational difficulties reviewed above were apparent. This is encouraging, because it points to the stability of the factors identified. Interestingly, a new difficulty surfaced while subjects went through condition (2), which involves the rule "if there is a vowel on the back of the card, there is an even number on the face of the card." To understand what is going on here, it is important to make the quantification over cards in the rule explicit: "for all cards, if there is a vowel on one side of the card, there is an even number on the other side of the card." A well-known issue in the semantics of the universal quantifier, first raised by Aristotle, is whether this quantifier has "existential import": if "all A are B" is true, does this entail that there *are* As which are Bs? Aristotle thought so, and he makes his point by means of two examples: of the two statements, "Some man is white" and "Some man is not white" one or the other must be true; and "Every pleasure is good" implies "Some pleasure is good." The examples show that Aristotle takes subjects and predicates to denote nonempty sets. This is precisely what our subjects appear to be doing, in requiring that, in order for the conditional to be true, there must be at least one card which satisfies both antecedent and consequent. They consider the rule to be undecided if there is no such instance, even in the absence of counterexamples.[24]

Subject 1. [in experiment (2)].
S. In this case I think you would need to turn the 7. And... that would be the only one you need to turn. ... 'Cause, these two [A and K] have to have numbers on the back, so they don't apply ... Which leaves these two [4 and 7]. Erm. And if there's an A on the back then it fits the rule [pointing to the 4] and if there is a K on the back then it doesn't apply to the rule, so it doesn't matter, which leaves this one [pointing to the 7], 'cause if there's an A on the back of here, and it's a 7, then you've disproved the rule.
E. OK, and if there is not an A on the back of that [the 7]?
S. If there's not an A on the back. [thinks] then ... maybe you do need to turn that one [pointing to the 4]. If there's not an A on the back [of the 7], then it doesn't disprove the rule and it doesn't prove it. So you'd have to turn the 4 I think.
E. And what if the 4 also didn't have an A on the back - what would that mean for the rule?
S. Well then ... for this set of cards, the rule... it would disprove the rule I suppose. Cause if there is not an A on the back of this card [pointing to the 4], [pause] then... there isn't a 4 on the face of it. OK, if there's not an A on the back, then none of these cards have an A on the back, and a 4 on the face, which is what the rule states. So for this set of cards it's disproved, it's not true.

It is abundantly clear that the subject considers existential import to be a necessary requirement for the truth of the conditional. The next subject expresses this insight by means of the modal "could" :

23. Performed in Edinburgh in 2003 by the our student Marian Counihan. The design was the same as that of the first experiment.
24. We will have more to say on this phenomenon in section 3.5.6 below.

Subject 7. [in experiment (2)].
S. If there's a A on the back of this card [the 7] then it's finished, you basically don't care anymore. Whereas if there's a K, all it seems to prove really is that this [the rule] could be true. ... I suppose we do... need to turn this card [the 4], just to affirm the rule.

The next subject has a slightly different interpretation, and says that if there is no card instantiating the rule (and no falsifying card), the rule is still undecided. This probably reflects a different understanding of what the domain of the rule is: the cards on the table, or a wider set of cards.

Subject 2. [in experiment (2): has stated he wants to turn 4 and 7]
E. OK, now say you turned both of those [pointing to 4 and 7] and you found an A on both.
S. Then the rule would be ... wrong. Because if there's an A on the back of both of them, then the rule says that there would be a 4 on the front of both of them, but, there's not, so, I mean there's a 4 on one of them, but then there is also a 7 on one of them, so, the rule's wrong, I mean, it doesn't always follow, so it's wrong. [Pause] Although it doesn't say always, there [pointing at the rule], but I am presuming it means always. I don't know [laughs].
E. Yes, it is meant to apply to any of them.
S. Yeah, OK, so the rule would be wrong if there's an A on both [4 and 7].

⋮

S. ... hang on, with an A and 7 there [on the 7 card], and an A and a 4 there [pointing to the 4] it would be wrong, still ... because of the A and 7, yeah, the A and the 4's correct, but because that [pointing to the 7 card] is incorrect, that's, the whole rule would be incorrect.
E. So, in either case, if there was an A on this side [pointing to the overturned 7 card], this other side of the 7, it would make the rule incorrect?
S. Yeah.
E. And despite what was on the [pointing to the 4]?
S. Oh, yeah, yeah. No OK, so you only need to turn that one [the 7]. Do you? [looks at rule] No you don't. No, sorry, no [indicates 4 and 7 cards again] you do need to turn them both, because if that is a K [turning over the 7], then you need to turn that one to check that one [the 4] to check that that is not a K as well.
E. And if that was a K as well?
S. Then the rule ... [pause] then you wouldn't know ... because ... there's nothing saying you can't have a K and a 4, but all it is saying is whether or not there's an A on the back, if there's a K on the back of both of these [the 4 and the 7] then you don't know, the rule might be right, or might be wrong.

These observations have some relevance to performance in the standard task. Suppose again, as we did in section 3.5.5, that subjects split the rule into the components (1) and (2). The first component yields the answer A; this card simultaneously establishes existential import. Interestingly, existential import as applied to the second condition yields the answers 4,7, so that we have found yet another way to justify the mysterious $p, q, \neg q$ response in the original task.

Interpreting the Conditional as a Conjunction

We now return to the possible interpretations of conditionals and their relevance to subjects' understanding of the task. In the literature on Wason's task only two types are distinguished: the unidirectional material implication, and the biconditional. When one turns to the linguistics literature, the picture is dramatically different. An interesting source here is Comrie's paper Conditionals: A Typology [44], where conditionals are distinguished according to the degree of hypotheticality of the antecedent. In principle this is a continuous scale. Viewed cross-linguistically, the degree of hypotheticality ranges from certain, a case where English uses *when* ("when he comes, we'll go out for dinner," [25] via neutral ("if a triangle is right-angled, it satisfies Pythagoras' theorem") to highly unlikely ("if we were to finish this paper on time, we could submit it to the proceedings") and even false, the counterfactual ("if we had finished this paper on time, we would have won the best paper prize"). If conditionals come with expectations concerning the degree of hypotheticality of the antecedent, this might affect the truth condition for the conditional that the subject implicitly applies. For example, we have seen in section 3.5.6 that some subjects claim the conditional has existential import; this may be viewed as the implicature that the antecedent of the conditional is highly likely.

Indeed we claim that, in order to understand performance in Wason's task, it is imperative to look into the possible understandings of the conditional that a subject might have, and for this, language typology appears to be indispensable. An interesting outcome of typological research is that the conditional ostensibly investigated in Wason's task, the hypothetical conditional, where one does not want to assert the truth of the antecedent, may not even be the most prevalent type of conditional. We include a brief discussion of the paper Typology of *if*-Clauses by Athanasiadou and Dirven [4] (cf. also [5]) to corroborate this point; afterward we will connect their analysis to our observations.

In a study of 300 instances of conditionals in the COBUILD corpus [42], the authors observed that there occurred two main types of conditionals, *course of event* conditionals, and *hypothetical* conditionals. The hypothetical conditionals are roughly the ones familiar from logic; an example is

> If there is no water in your radiator, your engine will overheat immediately. [42,17]

A characteristic feature of hypothetical conditionals is the events referred to in antecedent and consequent are seen as hypothetical, and the speaker can make use of a whole scale of marked and unmarked attitudes to distance herself from claims concerning likelihood of occurrence. The presence of "your" is what makes the interpretation more likely to be hypothetical: the antecedent need not ever be true for "your" car. Furthermore, in paradigmatic cases (temporal

25. Dutch, however, can also use the conditionals marker "als" here.

and causal conditionals) antecedent and consequent are seen as consecutive. By contrast in course–of– event conditionals such as

> If students come on Fridays, they get oral practice in Quechua (from [44])

or

> If there is a drought at this time, as so often happens in central Australia, the fertilised egg in the uterus still remains dormant [42,43]

the events referred to in antecedent and consequent are considered to be generally or occasionally recurring, and they may be simultaneous. Generic expressions such as "on Fridays" or "as so often happens ..." tend to force this reading of the conditional. e.g., the first example invokes a scenario in which some students do come on Fridays and some don't, but the ones who do get oral practice in Quechua. The generic expression "on Fridays," together with implicit assumptions about student timetables and syllabuses, causes the sentence to have the habitual "whenever" reading. It is also entailed that some students do come on Fridays, generally. These examples also indicate that course–of–event conditionals refer to events situated in real time, unlike hypothetical conditionals. It should now be apparent that the logical properties of course–of–event conditionals are very different from their hypothetical relatives. For example, what is immediately relevant to our concerns is that course–of–event conditionals refer to a population of cases, whereas hypothetical conditionals may refer to a single case; this *is* relevant, because it has frequently been claimed that subjects interpret the task so that the rule refers to a population of which the four cards shown are only a sample (cf. section 3.5.2). Interestingly, Athanasiadou and Dirven estimated that about 44% of conditionals in COBUILD are of the course–of–events variety, as opposed to 37% of the hypothetical variety. Needless to say, these figures should be interpreted with caution, but they lend some plausibility to the claim that subjects may come to the task with a nonintended, yet perfectly viable, understanding of the conditional. We will now discuss the repercussions of this understanding for subjects' card selections.

One of the questions in the paraphrase task asked subjects to determine which of four statements follow from the rule "Every card which has a vowel on one side has an even number on the other side." More than half of our subjects chose the possibility "It is the case that there is a vowel on one side and an even number on the other side." Fillenbaum [82] already observed that there are high frequencies for conjunctive paraphrases for positive conditional threats ("if you do this I'll break your arm" becomes "do this and I'll break your arm") (35%), positive conditional promises ("if you do this you'll get a chocolate" becomes "do this and I'll get you a chocolate") (40%) and negative conditional promises ("if you don't cry I'll get you an ice cream" becomes "don't cry and I'll get you an ice cream") (50%). However, he did not observe conjunctive paraphrases for contingent universals (where there is no intrinsic connection between antecedent and consequent) or even law-like universals. Clearly, the statements we provided are contingent universals, so Filenbaum's observations on promises and threats are of no direct relevance. However, if the course–of–event conditional is a possible reading of the conditional, the inference to a conjunction observed in many of our subjects makes much more sense. Clearly the truth–conditions for conditionals of this type differ from the intended interpretation; to mention but one difficult case, when is a generic false? Thus, a generic interpretation may lead to different evaluations and selections. 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 (since the K and 7 cards falsify); 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 actually caused by the conditional, and how much is caused by the task setting, no matter what binary logical connective is used?

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 generalization. The sociopsychological 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. 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. What would happen, for example, if the rule were stated using the putatively least problematical connective, conjunction? Chapter 4 gives the answer.

3.6 Matching Bias: the "No-Processing" Explanation

We have paid scant attention to more traditional interpretations of the selection task, focusing instead on the logical difficulties experienced by subjects in the descriptive task. This is not to imply that the traditional explanations are completely without foundation. We provide one example here: Evans's "matching strategy." This was proposed as a shallow processing strategy, operating automatically. We will have more to say on the distinction between shallow and deep processing in chapter 7 on the suppression task; but for now we give some examples showing that this type of response also occurs when subjects fully engage with the task.

Evans (see, for example, the review in Evans, Newstead, and Byrne [76]) defines the "matching strategy" as the choice of cards which match the atomic parts of the content of a clause in a rule. So for the rule *If p then q*, p and q cards match: for the rule *If p then not q* still p and q cards match: and the same for *If not p then q*. Here is a particularly striking example.

> *Subject* 9. [experiment (1)]
> *E.* [This rule] says that if there is a vowel on the face, then there is an even number on the back. So what we mean by face is the bit you can see, and by back the bit you can't see. Which cards would you need to turn over to check if the rule holds?
> *S.* This one [ticks A] and this one [ticks 7]
> *E.* So why would you pick those two?
> *S.* One has a vowel on the face and the other one an even number. If you turn it, if it's true, then it should have an even number [pointing to the A] and this should have a vowel [pointing to the 7].
> *E.* [baffled] So you picked, Oh you were saying if there was a vowel underneath [pointing to the 7]
> *S.* That's because I'm stupid. Even number is 1,3,5, ...
> *E.* No, 2,4,6, ...
> *S.* [Corrects 7 to 4, so her final choice was A and 4] OK. So these.

The next example is straightforward:

> *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.

Evans conceptualises the use of this strategy as a "superficial" response to both rule and task which subjects adopt prior to processing the information to the level of a coherent interpretation of the whole sentence. As such, the strategy may be applied prior to or alongside other processing strategies. It is taken to explain the modal response of turning the p and q cards in the abstract task. It must assume that something else is going on (perhaps superimposed on matching) when subjects adopt other responses. Thematic effects have to be explained in terms of contentful processes engaging other processes at deeper levels than matching.[26]

3.7 The Subject's Predicament

So what is the upshot of this extended semantic analysis of the range of subjects' interpretations and factors influencing them which were revealed by these Socratic dialogues? We feel the need to provide some more synoptic integration of this mass of rich observations, though what we offer here should be understood as a very partial view.

Some effects are global. For example, content may strongly shift subjects toward a deontic interpretation, and then few interpretational problems arise. The two cards Wason expected to be turned are the only possible violators of the "law." To produce an integrated sketch of some of these global effects, table 3.2 presents some of the global parameters an interpretation must fix, and hints at their relations. Under descriptive interpretations, robust ones (tolerating exceptions) lead immediately to conflict with finite sets of cases determining truth–value, and problems of distinguishing exceptions from counterexamples, which in turn may lead to the "cards as sampled from a population" interpretation. Brittle interpretation allays these problems but raises the issue of contingencies between card choices. Both robust and brittle interpretations are susceptible to both kinds of reversibility – of physical cards, and of logical rule. Deontic interpretations suffer from none of these problems: cards are independent of each other, only cases are judged (not rule), the veracity of the experimenter is not at stake. The content triggering a deontic reading generally makes for logical irreversibility, and there are no anaphors to interact with card reversibility issues.

But many of the factors affecting interpretation are local and interact with other parameters in determining the interpretive outcome, defying tabulation. A useful supplementary way to draw together the complex threads is to tabulate some ranges of interpretations which can lead to each of the four dominant choice combinations in table 3.1, which jointly account for 92% of the subjects. Table 3.3 lays out some of the parameters leading to these common choice combinations.

26. Oaksford and Stenning [204] by investigating a full range of clause negations in both selection and evaluation tasks, showed that matching is not a particularly good explanation of performance with the full range of negated conditionals. They argue that a better summary of the data is in terms of the degree to which the material and instructions allow negative clauses to be processed as corresponding positive characterizations.

Table 3.2 Global parameters as they contrast between descriptive and deontic interpretation

<table>
<tr><td colspan="3">descriptive</td><td>deontic</td></tr>
<tr><td colspan="2">robust</td><td>brittle</td><td></td></tr>
<tr><td colspan="2">conflict with task
“population” reading</td><td>dependencies between cards</td><td>independent cards
judging cases–not rule</td></tr>
<tr><td colspan="3">reversibility</td><td>no reversibility</td></tr>
<tr><td>of cards</td><td colspan="2">of rule</td><td></td></tr>
<tr><td colspan="3">discomfort with challenging E</td><td>no calling E a liar</td></tr>
</table>

Table 3.3 Some interpretation features and the main choice-combinations to which they may lead in the descriptive selection task: some choices on other tasks are mentioned with their interpretational feature.

Main card choice combinations				Interpretational Feature
p	p, q	$p, \neg q$	$p, q, \neg q$	(in italics)
x	x			*strong negation*: choose either (both) letter(s), or $\neg q$ in two–rule task
			x	*not-false doesn't mean true*: ruling out counterexamples, but also seeking positive case
		x		*truth vs. satisfaction ambiguity*: if resolved, removes contingency issues if unresolved in two-rule task, reject materials
	x			*robustness*: immediate conflict with task, may invoke sample reading
x				*choice contingencies*: choose true antecedent first RAST and two-rule tend to remove problem
	x			*can't call E a liar!*: assume rule true, seek cases that make rule true
	x			*in conjunction task*: test cards of unknown truth–value
			x	*existential import of universal* rule out counterexamples but positive instance required too
x	x			*assume cards irreversible*: maintain logical irreversibility, or convert and conjoin conditional
	x			*superficial mention in rule* determines relevance "matching"
			x	*deontic conjunctive interpretation* or *one-card-at-a-time descriptive interpretation*

3.8 Conclusion

The explorations by dialogue reported here have cast their net somewhat wider than is customary, to obtain information about subjects' processing and semantic understanding of the task. The picture that emerges is complex. The differences between subjects, even when they make the same selection, are huge and defy any single explanation. All choice combinations reflect more than one interpretation and some combinations we are still struggling to explain. For example, there appears to be no explanation for a very common pattern of evaluation and selection: p, q is selected, $q/\neg p$ is evaluated as falsifying, and $\neg p/q$ is evaluated as irrelevant, although we ventured a hypothesis in section 3.5.5. It does not seem very helpful to dismiss such behavior as "irrational." It is more interesting to relate this and other behavior to subjects' understanding of the task, but much richer data are required than the four "bits" received from each subject in the classical task. We have seen that understanding interpretation sometimes leads to clarification of what subjects are trying to do, and that often turns out to be quite different than the experimenter assumes.

The granularity of these data is very much finer than experimental psychologists have deemed necessary for the analysis of this task (though this fine granularity is rather ordinary in lots of areas of psychology such as visual perception). A common reaction is that this interpretational variety is all very well for linguistics, but surely it is obvious that this is not what is going on in the few seconds that subjects think about what cards to turn over in the original task, and besides, if this is what is going on, it's far too hard to study. This attitude was crystallized for KS by a remark made after he had given a talk on the material (ironically in Wason's old department). "You have given a very interesting analysis of the task from a semantic point of view," the questioner commented, "but what we need is a psychological explanation." The questioner seemed to assume that when linguists and logicians substantiate an analysis, that they are claiming "ordinary speakers" have conscious access to the analysis within its own terminology. Equally, they assume that even demonstrating the formal semantics is correct doesn't have the implication that ordinary speakers actually mean by the expression, what the analysis says they mean.

We would be the first to acknowledge that our analysis isn't more than a beginning, but what is striking here is the idea that it is not a "psychological" analysis. Cognitive psychology is founded on the idea that people interpret their fresh experience in the light of their long–term knowledge and that the resulting rich structures furnish them the wherewithal to reason to their decisions. If we don't know how subjects interpret a task, or its materials, then how are we to start understanding what they do? And if they do lots of different things which are each comprehensible on the basis of their different interpretations, then hadn't we better have different explanations for their different acts? In some subfields of psychology, the questioner's comment might have been taken to mean that what is needed is a *mental process* model and that that is what sets semantics off from psychology. We have some sympathy since we agree that process models would be really nice to have, and we certainly don't have one, although the logical models in section 3.4 are an essential prerequisite for a process model. None of the theories of this task are any more process theories than the approach offered here – most of them less so. And if reasoning to an interpretation among a rich set of possibilities is the mental process going on, then how can we get a process model without any account of the range of interpretations in play? In the next chapter we will bring evidence that the gap between the interpretational problems appearing in these dialogues and what goes on in the traditional controlled experiments is not so great as it might at first appear.

These dialogues may also provide challenges to natural language semantics. While it is of course possible to attribute the vacillations in interpretation (of conditionals and anaphora, for example) to performance factors, it seems more interesting to look into the structure of the linguistic competence model to see how the observed interferences may arise. It seems to us that dialogues such as these provide a rich source of data for semantics and pragmatics, which promises to yield deeper insight into interpretation and processing of natural language.

What we hope to have demonstrated in this chapter is that the data do not warrant abandoning the search for formal models to provide bases for explaining subjects' reasoning behavior. Instead, formal models embodying insights from neighboring fields are useful guides for a richer program of empirical exploration and testing. There is a danger that deceptively simple models obscure the phenomena in need of explanation, and in so doing likewise obscure the educational relevance of the logical competence models and their highly objectified stance toward language. Stanovich [254] shows how closely related this stance is to other educational achievements. The tutorial dialogues presented here provide some insight into the variety of students' problems which may be of some help to those involved in teaching reasoning skills.