HAL
open science

Reasoning with conditionals

Guy Politzer

To cite this version:

Guy Politzer. Reasoning with conditionals. Topoi, 2007, 26 (1), pp.79-95. 10.1007/s11245-006-9001-9 . ijn_00184577

HAL Id: ijn_00184577

https://hal.science/ijn_00184577v1

Submitted on 31 Oct 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Published in Topoi, 2007, *26*(1), 79-95.
Original publication available at www.springerlink.com

Reasoning with conditionals

Guy Politzer

Abstract

This paper reviews the psychological investigation of reasoning with conditionals, putting an emphasis on recent work. In the first part, a few methodological remarks are presented. In the second part, the main theories of deductive reasoning (mental rules, mental models, and the probabilistic approach) are considered in turn; their content is summarised and the semantics they assume for *if* and the way they explain formal conditional reasoning are discussed, in particular in the light of experimental work on the probability of conditionals. The last part presents the recent shift of interest towards the study of conditional reasoning in context, that is, with large knowledge bases and uncertain premises.

Introduction

Conditional if-statements are pervasive. Their ubiquity and their occurrence in the development as early as the end of the third year (Scholnick and Wing, 1991, 1992) testify to their essential role in language and cognition. Not surprisingly, their investigation in linguistics, philosophical logic and psychology is very lively. For many years, psychologists were developing their own research in isolation, but this is no longer the case. In this paper, psychological investigations of reasoning with conditionals will be reviewed and connected with work in other disciplines. The first part contains methodological considerations that help understand psychological research in the area of reasoning with conditionals. The second part develops the main theoretical approaches that aim to account for formal deductive reasoning together with recent experimental studies on the semantics of *if.* The third part describes recent investigations of conditional reasoning that are more respectful of ecological validity, that is, using statements whose content refers to reasoners' knowledge base.

1. The investigation of reasoning with conditionals: methodological issues

Besides a number of inferences such as contraposition, transitivity of *if,* etc., psychological investigations of reasoning with conditionals exploit three tasks that date back to the late 1950's or the early 1960's. They are: the truth table task, the so-called "conditional reasoning task", and Wason's four-card problem (also known as the Selection Task).

The first one, the truth table task, has been little used as compared to the other two. It was developed to compare the interpretation of conditional statements by naive reasoners (that is, individuals who have no academic background in logic) with the material conditional which was the almost unique reference at the time, especially among psychologists. The participant is presented with a conditional statement *if A then C* and then with the four logical cases in turn *(A, C; A, not-C* etc.). Depending on the materials, the presentation of the logical cases may be verbal, in the form of a drawing, or of concrete objects, etc. The participant is asked to <u>construct</u> the cases that make the

statement true and that make it false, or to evaluate the statement as true or false. In the construction procedure (used by Evans, 1972) the combinations overwhelmingly chosen are *A, C* to make the statement true and *A, not-C* to make it false while the *not-A, C* and *not-A, not-C* cases are seldom chosen. This is observed when participants are asked to choose only one combination (Evans, Legrenzi and Girotto, 1999; Oaksford and Stenning, 1992) or when this number is left free (Evans, 1996). In sum, reasoners seem to consider the *not-A* case as irrelevant to the truth value of conditional statements. This is confirmed by the evaluation procedure which yields an important proportion of reasoners who declare the *not-A* case to be irrelevant; more details will be given later.

There is another observation that deserves a comment right now. One of the most robust results is the existence of an interpretation in which the statement is evaluated as true in the *A, C* and *not-A, not-C* cases, and false otherwise, that is, as a biconditional. Independently of this observation, in a seminal paper, Geis and Zwicky (1971) pointed out that certain conditional statements, such as contractual promises (e. g., *if you mow the lawn, I'll give you five dollars)* are usually understood as implying their obverse *(If you don't mow the lawn, I will not give you five dollars)* so that the final logical form reduces to a biconditional. They called the implied sentence an "invited inference". The wide acceptance of the obverses of promises, threats, and causals, has been experimentally observed by Fillenbaum (1975, 1976) and regularly confirmed by subsequent work; he concluded that the illocutionary point of a statement may affect or control its logical form. This kind of implied proposition is generally referred to as a "conversational implicature", following Grice's (1975) terminology. *If* is not the only term that can trigger an implicature; other logical terms (connectives and quantifiers) have been investigated as a result of the development of pragmatic theory (Gazdar, 1979; Horn, 1989; Levinson, 2000; Noveck and Sperber, 2004). Not all theorists agree on the mechanism that leads to the implied meaning, but we are not concerned with this debate in this review. Suffice it to emphasise that accepting the obverse (or the converse) of a conditional need not be a logical fallacy for it may be the legitimate result of a pragmatically implied inference. One related point of paramount importance for the psychology of reasoning is that the identification of the logical forms of natural language sentences that are used for observation purposes or for the construction of experimental material is a difficult technical enterprise. This requires a semantic and pragmatic, as well

as a logical, scrutiny of the sentences that constitute the arguments under study in order to ensure the content validity of the task, that is, that the participants process the logical form that is intended by the experimenter (Politzer and Macchi, 2000; Politzer, 2004). The necessity of such analyses is reinforced by the fact that in experimental situations participants attribute intentions to the experimenter, and a condition for them to answer successfully is that this attribution process be correct. Now, the more the ambiguities to resolve, the worse are the chances to get at the right intentions.

The second task consists of four arguments, each with the same conditional premise, presented in turn. Two are deductively valid; they are Modus Ponens (MP) and Modus Tollens (MT) and the other two are the fallacies of Affirming the Consequent (AC) and Denying the Antecedent (DA). When the four arguments are presented as a whole set, the task is referred to as "the conditional reasoning task".

We now introduce some of the psychologists' jargon and experimental procedures. In order to study human capability to reason formally, the experimental material is chosen in such a way that the context should be as neutral as possible. One solution is to describe microworlds made of items such as letters, geometrical figures, etc., sometimes characterised by various features (size, colour, etc.) and use sentences that express facts or properties that are clearly arbitrary (e. g., *the cube is blue or small,* or *if there is a square, there is not a circle,* etc.) One may opt for more abstract material and use sentences without any context such as *if P then Q* (but with a risk that some participants spontaneously instantiate the sentences for themselves). In all these cases, the materials are called "abstract". In contrast, when a context is offered, either explicitly after the description of a scenario, or implicitly through the mere content of the atomic sentences, the materials are called "thematic".

It is also important to bear in mind the existence of two main procedures. In the "production" procedure, only the premises of the argument are presented, and the participants are instructed that they have to either produce a conclusion, or indicate that nothing follows. In the "evaluation" procedure, a whole argument is presented and participants are asked to evaluate its conclusion by choosing among options such as *the conclusion is true, the conclusion is false,* and *the conclusion may be true or false.*

Usually, the question put to the participants is to produce a conclusion that "follows logically", or to decide whether the conclusion given "follows logically". This is

not without problem because the notion of logical consequence may not be intuitively obvious for participants. The overall percentages of individuals who endorse the arguments may be overestimated, as "following logically" might turn out to be a catch-up qualification capturing various kinds of consequences: probabilistic, plausible, in addition to strictly logical.

With these reservations in mind, results of meta-analyses (limited to a conditional premise without negated antecedent or consequent) indicate that MP is endorsed by 95 to 100 percent of the individuals, MT by about two thirds, and AC and DA by about half of them. Explaining this pattern of response is one of the main objectives of any psychological theory of conditional reasoning. These arguments could even be considered as benchmark problems for any theory of reasoning at large.

The last task, the selection task, probably is the most famous of the tasks devoted to the study of human reasoning, especially outside the psychology community. It would be more appropriate to use a plural, for there are two versions of the task, which have only superficial resemblance. One is thematic and has been investigated mostly in the deontic domain; it will not be discussed for lack of space. The other version is formal; it was invented by Wason (1966) and has intrigued many philosophers and economists interested in human rationality. Participants are presented with the picture of four cards. They are instructed that each card has a letter on one side and a number on the other side. Two of the cards show the letter side up and the other two show the number side up, such as: [A] [K] [3] [7]. The question is: "Indicate those cards and only those cards that need to be turned over in order to decide whether the following rule is true: *If there is an A on one side, then there is a 3 on the other side"*. Even though this task has been the object of innumerable studies, it will not be considered in this review because after forty years of investigation (and contrary to the early times), there is no agreement about its nature (e. g., in terms of deduction versus induction, or of reasoning versus decision making), on the possibility for participants to represent the task unambiguously (Stenning and van Lambalgen, 2004), nor is there agreement on a normative model and consequently on what the correct solution is (even though the vast majority defines the correct solution as Wason did, that is, selecting [A] and [7], which participants seldom do). For lack of validity (it is not clear what reasoning skill is involved) the task is not even adequate for testing between theories of

reasoning. A defence of this negative view, both theoretical and empirical, is offered by Sperber, Cara and Girotto (1995). Ironically, the formal task might be more interesting from the viewpoint of the sociology of science, and of what it reveals about the evolution of psychologists' view on the logic of conditionals and also on induction. The task was born in the background of Popper's falsificationism and it required about a quarter of a century for a different normative approach to appear, namely the probabilistic approach (Oaksford and Chater, 1994), which vindicates the rationality of the most frequent selection after the [A] card, viz. the [3] card. It follows, incidentally, that the doubts against human rationality aroused by the performance on the abstract task must be established on different and better evidence. A thorough discussion of the Selection Task can be found in Poletiek (2001).

2. The main approaches to reasoning with conditionals

We now consider in turn (sections 2.1 and 2.2) the two main theories of deductive reasoning. We summarise the principles on which they are based and review the semantics they assume for the conditional, and then how they explain performance on the tasks and arguments of interest. Contrasting with these two logic-based approaches, we present briefly a probabilistic approach to conditional reasoning (2.3). This is followed by a summary of recent experiments on reasoners' conception of the probability of conditional statements (2.4).

2.1. Mental rules

Of the two main models of deductive reasoning, Rips's (1994) Psycop, and Braine's (Braine, Reiser, and Rumain, 1984; Braine and O'Brien, 1998) "mental logic", only one will be considered, if only because their inspiration is very similar. We consider the latter because it offers a comprehensive theory of the conditional. We begin with a short description of the model which is indispensable to understand the treatment of the conditional.

2.1.1. An overview of the mental logic theory of deduction

The model follows a syntactic approach to human deductive inference. It assumes the existence of a universal "syntax of thought" that acts on semantic representations, with which it constitutes a system akin to Fodor's language of thought.

The model does not include comprehension mechanisms; that is, it leaves out the processes by which the surface structure of natural language propositions is transformed into semantic representations. The model comprises two components. One is a set of inference schemas that constitute a natural logic (in the sense of Gentzen). The other is a reasoning program that selects schemas to be applied to the premises to construct the shortest possible "chain of reasoning" leading to the conclusion. This program has two parts. The first part, the "direct routine", is always executed, whereas the second part, the "indirect strategies" takes over in case the first has failed to evaluate or find a conclusion. The inference schemas and the direct routine are assumed to be universal, early acquired skills, independent of language, culture and education, which is not the case for the indirect strategies which are later acquisitions that can be affected by factors such as education. Its use requires heuristics to find a successful chain of reasoning.

The inference schemas do not all have the same status. There is a core of 7 schemas such as double negation and MP which can be applied routinely. The remaining 6 are submitted to specific constraints; this group contains and-elimination, and-introduction, conditional proof (detailed later), reductio ad absurdum, and the following two schemas:

p ; NOT-p / INCOMPATIBLE

p OR q ; NOT-p AND NOT-q / INCOMPATIBLE

The last four have a restricted role in the direct program. The last two define contradiction: the expression "incompatible" reflects a psychological assumption to the effect that reasoners react to a contradiction by estimating that no conclusion -- rather than anything -- follows.

The direct routine of the reasoning program has three parts (the first and the third of which apply only when the conclusion is presented together with the premises). The first part concerns only conditional conclusions (about which more later). The third part is an evaluation procedure of the given conclusion. Essentially, it makes use of four of the schemas to test for identity or contradiction with the current set of premises and responds *true* or *false*. The second part of the direct routine is the inference procedure

proper. A core schema is applied when the current premise set contains a proposition whose form matches the "numerator" of the schema immediately or after first applying double negation, or and-elimination, or and-introduction. Then the "denominator" of the schema is added to the premise set to constitute the current set. When there is a conclusion to be evaluated, the evaluation procedure is used (part 3). If the outcome is indeterminate, the inference procedure is repeated until the conclusion is evaluated. The direct routine terminates when the conclusion is evaluated or when the inference procedure does not generate new propositions. In the latter case the indirect strategies are considered (and the direct routine will be re-activated).

As far as the indirect reasoning program is concerned, four strategies are described (without a claim for exhaustivity), such as reductio ad absurdum and the use of a lemma. The capability to use such strategies is basically what distinguishes good and poor reasoners, since the schemas and the direct program are common to all reasoners. It follows that arguments that do not require the application of strategies should not give rise to differences in performance among reasoners, provided there is no overload in working memory. When the number of inference steps in a deduction increases, the difficulty of the argument increases (whether it is objectively defined by the error rate or by participants' report of mental effort); individual differences may then arise.

The model has been evaluated (with arguments that do not involve indirect strategies) in a series of experiments using formal materials (Braine and O'Brien, 1998, chapters 7 and 8). The results offer a convincing demonstration of its adequacy at least within the constraints chosen for the experiments. A discussion of the results is beyond the scope of this paper.

2.1.2. The mental logic theory of *if.*

Braine and O'Brien's (1991, henceforth BO'B) theory of *if* has three components. The first concerns the semantics proper, called "the lexical entry", which contains the information kept in the individual's semantic memory. The second component contains pragmatic principles that, together with the lexical entry, lead to the interpretation of sentences in context. The third component contains the propositional reasoning program described earlier. The pragmatic principles need not be described here

because they are not specific to the theory; they are consistent with the major pragmatic theories (e. g., relevance theory, Sperber and Wilson, 1986/1995) that aim to account for the processes by which the hearer, starting from the linguistic meaning of a sentence, builds an interpretation that ultimately yields the semantic representation, on which the reasoning program acts.

The lexical entry is defined by the inference schemas of the reasoning program that involve *if.* The schemas are viewed as providing instructions about the inheritance of truth from premises to conclusion. In brief, the model adopts a procedural semantics. The lexical entry is primarily defined by the two inference schemas which involve *if* alone, namely Modus Ponens and the Schema for Conditional Proof :

(MP): Given *if p then q* and *p,* one can infer *q,* and

(CP): To derive *if p, then...,* first suppose *p;* for any proposition *q* that follows from the supposition of *p* together with other information assumed, one may assert *if p, then q*.

The CP obeys a special constraint that is essential to the theory, as it restricts the arguments allowable in the system. The constraint stems from the postulate that for ordinary reasoners nothing follows from a contradiction, except the notion that some initial assumption is wrong. Consequently, for the CP to function, the supposition must be consistent with prior assumptions (but the reasoner may not always know in advance whether this is the case). There follows a corrolary of great practical importance: a premise reiterated into a conditional argument cannot contradict the supposition, and this is the only consistency with which the reasoner need usually be concerned (for any non reiterated premise that would be inconsistent with the supposition can be discounted as superfluous).The computational difficulty of the search for consistency is greatly alleviated by this limitation; another factor is the relatively small size of the premise set that is allowed by human processing.

The operation of the constraint has a number of important consequences. One, contraposition generally is valid (its proof involves a supposition and a *reductio).* However, there are circumstances where it is invalid: These arise whenever the constraint is activated, e. g., by a semantic relation between *p* and *q* (which amounts to adding an assumption) or, even more simply, by the truth value of *q*: when *q* is true anyway (cf Austin's "biscuit conditionals" such as *if you are hungry there are biscuits in*

the cupboard) contraposition fails. Similarly, transitivity is valid, except when it is blocked by the constraint for reasons that pertain to the relations between the three atomic propositions (for an example, see Braine and O'Brien pp. 186-187). For these reasons, contraposition and transitivity cannot be formulated validly by a general derived schema.

Two, another property of the model is the way it handles the refutation of ordinary conditional statements. This is taken in charge by the reasoning program which stipulates, for arguments whose conclusion is a conditional, to add the antecedent to the premise set and treat the consequent as the conclusion. (This is the first part of the direct routine mentioned earlier). By CP, finding *not q* allows to infer *if p, then not q,* but because *p* false is excluded by the constraint, the inference contradicts *if p, then q*. In brief, the system nicely captures the natural way of negating a conditional, by just negating its consequent.

Three, the truth of a conditional statement does not follow from the falsity of the antecedent because the premise and the supposition are inconsistent and the constraint blocks the inference. This settles one of the two paradoxes of implication. The other paradox coincides with a valid inference in the system and therefore requires an explanation. It is the common, pragmatic, explanation that is offered: if *q* is known to be true anyway, there is no point in asserting *q* conditioned on *p.*

Lastly, the CP provides a procedure to evaluate the truth of a conditional in a given state of affairs (which provides a set of premises): one supposes the antecedent true in that state of affairs and one tries to derive the consequent; if one succeeds, the conditional is evaluated as true and if one derives the negation of the consequent, the conditional is evaluated as false. The initial step of the procedure clearly is a version of the Ramsey (1931) test (in line with the psychological interpretation and exploitation of it proposed by Evans and Over, 2004). This explains performance on the truth table task: when the antecedent is presented as true the evaluation procedure applies and yields *true* or *false* for the conditional depending on whether the conditional is true or false (because it can be reiterated as such in the suppositional argument). But when the antecedent is presented as false, the CP cannot be applied (unless the state of affairs could be changed) and no truth evaluation is possible. This explains the defective truth table and, of course, reasoners' intuitions.

Performance on the four basic arguments is explained as follows. MP belongs to the lexical entry of *if* and as such belongs to the core schemas of the reasoning program. This explains its universal endorsement. MT requires a strategic use of *reductio ad absurdum,* which not all reasoners possess, hence a moderate performance. The two invalid arguments (AC and DA) are endorsed for interpretative reasons. A proportion of participants routinely add an invited inference which turns the conditional premise into a biconditional with which the arguments are valid. This explanation is supported by strong experimental evidence. One procedure consists of adding a third premise in order to countermand the biconditional interpretation. (Experiments using thematic materials will be described later). One experiment using arbitrary material was run by Rumain, Connel and Braine (1983). A control group was presented with a major premise such as *if there is a dog in the box, then there is an orange on the box* together with an appropriate minor premise (e;g., for AC, *there is an orange in the box)* while an experimental group was presented with the same premises and the additional premise *if there is a tiger in the box, then there is an orange in the box*. Averaged over children and adults, the rate of endorsement of both arguments dropped from about three quarters of the participants to one third. While the decrease is impressive, it remains to be explained why one third of participants still commit the fallacy; more on this in section 3. 2.

There are similarities between the BO'B theory of *if* and Stalnaker's conditional supplemented with the Stalnaker-Thomason (1970; Thomason, 1970) logic. One might view the former as a variant of the latter with alterations that aim to make it psychologically more plausible. It drops the evaluation of truth values in terms of possible worlds, and especially the use of a selection function. It also departs from the Stalnaker-Thomason logic in that it has one single conditional, eliminating the material conditional altogether. In their discussion of this logic, BO' B point out that this system allows many inferences which ordinary reasoners do not make and also that it disallows many others which thay make (due, in particular, to the operation of the reiteration rules).

To conclude on the BO'B theory of *if,* it offers solutions to the main problems raised by the psychology of conditional reasoning, except that in its present state it is silent on the question of the probability of conditionals, which does not allow it to apply to uncertain reasoning. But there is nothing in principle that prevents a definition

of the probability of a conditional: it suffices to apply the CP schema along with the truth evaluation procedure; because the antecedent is supposed to be true, the false antecedent cases do not intervene in the evaluation of the conditional and the outcome coincides with the conditional probability.

2. 2. Mental models

2. 2. 1. An overview of the mental models theory of deduction

The mental models theory of deduction developed by Johnson-Laird and collaborators (Johnson-Laird, 1999, 2001; Johnson-Laird and Byrne, 1991) is a semantic -- as opposed to syntactic-- theory. Mental models are representations of situations based on the understanding of discourse (in particular, premises) perception or imagination. "Each mental model represents a possibility and its structure and content capture what is common to the different ways in which the possibility might occur" (Johnson-Laird, 1999). Although a model is iconic, (that is, its parts and structure correspond to the parts and structure of the possibility), they are distinct from images as they can contain abstract elements and represent situations that cannot be visualised. Models can represent situations, events, processes; in particular they can represent propositional connectives. For instance, a statement about some figures on a blackboard such as *there is a circle and there is a star* can be represented by: O *

The construction of models follows two principles: (i) each model represents a possibility; (ii) mental models represent what is true, but by default not what is false (principle of truth).Take for instance *there is a circle or there is a star but not both.* There are two true possibilities, each of them represented by a distinct model on a separate line:

O

*

It is false that there is a star in the first model and it is false that there is a circle in the second model but this is only implicitly represented by the abscence of the corresponding figures. People are assumed to keep this information in the form of a "mental footnote". But the models can be, if necessary, fully explicit as follows:

O ¬*

¬O *

Not only parts of models can have an implicit representation but also whole models, as will be seen shortly.

Deductive reasoning consists in the manipulation of models. Given a set of premises, there are a few principles of combination of the premises' models to arrive at integrated models of the premise set. Then, deduction is based on the principle that a putative conclusion follows from the premises if it holds in all the models of the premises (i. e., if it has no counterexamples). If it holds in at least one model, the conclusion is possible. The main general prediction of the theory is that inferences that require more models lead to more errors because of the increased load in working memory. Also, for the same reasons, individuals tend to forget the implicit elements in the models and to focus on what is true and neglect what is false.

The theory of mental models has been applied to, and tested in, a variety of domains other than propositional reasoning, e. g., syllogistic (Johnson-Laird and Bara, 1984), relational (Byrne and Johnson-Laird, 1989), modal (Bell and Johnson-Laird, 1998), and certainly more widely than mental rules, which are limited to propositional and predicate reasoning. Many of its predictions tested by its proponents have been experimentally supported but the results are, contrary to the claim, often susceptible of receiving alternative explanations in terms of mental rules. At present, it is by far the most influential theory of reasoning.

2. 2. 2. The mental models theory of *if*

Johnson-Laird and Byrne (2002) (henceforth JL&B) introduce a distinction between a theory of comprehension which relies on models, and a theory of meaning which relies on possibilities. Possibilities can be either factual or counterfactual (i. e., false at the time of the utterance). Another distinction concerns the claims made about a situation which can be either factual or modal.

The meaning of conditionals (the "core semantics") is given for the "basic" conditionals which are defined as "those with a neutral content that is as independent as possible from context and background knowledge, and which have an antecedent and

consequent that are semantically independent..." The antecedent of a conditional, *if A...* establishes two possibilities: either two factual possibilities (a, ¬a) where *a* denotes the possibility satisfying the antecedent, and ¬*a* the possibility satisfying its negation; or a factual and a counterfactual possibility (the counterfactual can be *a* or ¬*a).*

There are two sorts of basic conditional (first principle):

- the core meaning of *if A then C* is defined by the factual possibilities:

a	c	
¬a	c	
¬a	¬c	(conditional interpretation)

- the core meaning of *if A then possibly C* is defined by the same factual possibilities augmented with [a ¬c], yielding the "tautological interpretation".

The conjunction of *if A then C* and *if not-A then not-C* (or *if C then A*) yields the factual possibilities:

a	c	
¬a	¬c	(biconditional interpretation)

The second principle concerns counterfactuals and will not be developed.

The third principle states how models are constructed, based on the foregoing sets of possibilities. It postulates that only the possibilities in which the antecedent is satisfied are explicitly represented; otherwise, they are implicitly represented. Hence the representation for *if A then C*:

a	c
- - -	

where the dots indicate the other possibilities as a single implicit model. Similarly, the representation of *if A then possibly C* is:

a	c
a	¬c

The last two principles state that the construction of models is constrained by (i) the meaning of, and coreferential links between, the antecedent and the consequent (semantic modulation); and (ii) the models representing contextual information which can modulate the core interpretation (pragmatic modulation). This can add information to models, prevent the construction of models, or help turn implicit models into explicit

models. An example of semantic modulation is given by *if it's a game then it's not soccer,* where the model

¬game soccer

is not constructed, leaving only two models:

game ¬soccer

¬game ¬soccer

An example of pragmatic modulation is given by *if a match is struck then it lights* and *the match is soaked in water and then struck.* The conditional is represented by:

match-struck - - - match-lights

and the categorical statement by:

match-soaked match-struck

The two premises are integrated to yield:

match-soaked match-struck match-lights

The knowledge base provides *if a match is soaked wet, it will not light* represented explicitly as:

match-soaked ¬match-lights

¬match-soaked ¬match-lights

¬match-soaked match-lights

Its combination with the model of the premises would yield a contradiction:

match-soaked match-struck match-lights ¬ match-lights

but precedence is given to general knowledge, hence:

match-soaked match-struck ¬match-lights

that is, *the match is soaked, struck and does not light.*

The operation of semantic and pragmatic modulation on the core interpretations *if A then C* and *if A then possibly C* can preclude some possibilities. This yields ten sets of possibilities (and this also apply to deontic conditionals). For lack of space, we only consider three of them for *if A then C:*

- relevance:

a c

¬a c

in which the consequent holds in any case. A typical example is Austin's "biscuit conditional".

- tollens:

¬a ¬c

which exploits knowledge of the consequent's falsity to convey the falsity of the antecedent. A typical example is Strawson's "hat conditional": *if he has passed his exam, I'll eat my hat.*

- ponens:

a c

which exploits knowledge that the antecedent is obviously true to convey the truth of the consequent.

We now examine how the theory accounts for the main phenomena.

Firstly, consider performance on the four arguments of conditional reasoning. Three levels of expertise are distinguished: elementary, where the footnotes attached to the implicit models are not even taken into account; intermediary, where footnotes indicating that A is false (for the conditional) and that A and C are false (for the biconditional) are taken into account; higher, where the full models are made explicit. To show how the theory functions, we focus on MP and MT. At the elementary level for MP, combining the model of the minor premise [a] with [a c] straightforwardly yields the conclusion C; but for MT, the combination of [¬c] with [a c] yields the null model and no conclusion follows. At the next levels, for MP whatever the interpretation -- conditional or biconditional-- the model of the minor premise eliminates the implicit model and the conclusion C follows from the model [a c]. On the contrary, for MT the minor premise *not-C* results in a model [¬c]; if the reasoner is focused on the model [a c], this suggests again that nothing follows. For MT to be executed correctly, the reasoner must construct the explicit model [¬a ¬c] from which the conclusion not-A follows. The theory explains performance in a similar manner, *mutatis mutandis,* for AC (but seems to run into difficulty for DA with the conditional interpretation at the intermediate level).

Considering now the classic problems of the conditional, we begin with how to negate a conditional. JL&B do not treat this question, but one may assume either one of two solutions. Given the models

a c

- - -

either reasoners cannot negate implicit models and no answer follows; or they make the models fully explicit and they take the complementary set, which yields [a ¬c] , that is the negation of the material conditional. Unfortunately, whichever the case, the result is wrong as normally reasoners answer *if A then not C,* which should be represented either by

a ¬c

- - -

or by the fully explicit models.

Next, the "paradoxes" of implication. As JL&B mention, the possibilities of the core meaning of *if A then C* correspond to the material implication. It follows that reasoners should accept the two inferences as non paradoxical. The authors explain their counterintuitive character by two factors. One, the judgment in terms of truth value concerns a meta-ability which JL&B regard as less appropriate than a judgment in terms of possibility; this explanation seems rather weak. The other factor is that the two inferences throw away semantic information, which means that there are more possibilities in the conclusion than in the premises. But it is easy to show that this condition is neither necessary nor sufficient. To see, for instance, that it is not sufficient, just consider that except when the "possibilities" in the conclusion of a deductive inference are the same as those in the premises (that is, when the inference expresses an identity) there are always more "possibilities" in the conclusion of a valid inference (because the "possibilities" are logical models). Luckily, not all valid inferences are odd. In all likelihood, what JL&B mean is inferences that throw away semantic information <u>in an intuitively obvious manner</u>, but this need a precise definition. We conclude that JL&B do not have a satisfactory explanation of the paradoxes of implication, which suggests that the conditional of ordinary language is not captured by the core meaning assumed by the authors.

Transitivity (and presumably contraposition) functions so long as the conclusion is believable. Otherwise individuals judge that nothing follows, but no detail is given as to how this response is generated.

To explain the judgment of irrelevance on the truth table task, JL&B assume that the two possibilities in which A is negative have implicit models which are not made explicit. This is fine for the production procedure. But in the evaluation procedure, participants are presented with the four logical cases, which means that all the models are materialised, in particular the implicit models are made explicit for the reasoner. Still, in response to a true/false question, the *not-A* cases are evaluated as *false* about half the time (Paris, 1973; Delval and Riviere, 1975); the latter study even mentions one third of spontaneous justifications in terms of irrelevance, an observation that suggests that JL&B's dismissal of the results of Johnson-Laird and Tagart (1969) on the ground that the *irrelevant* option was offered to the participants might not be justified. Politzer (1981) observed about one half of *incompatible* answers to a *compatible/ incompatible* response format for the not-A case, with frequent justifications in terms of irrelevance. This result is interesting because the wording of the question does not fall under JL&B's objection that true/false questions tap a metalinguistic ability. In brief, the MM theory of *if* can explain performance on the truth table task for one procedure but fails for the other.

The foregoing invites one to ask the question of the truth-functionality of *if* in JL&B's theory. The authors claim most explicitly that it is not. But they seem to have a special notion of truth-functionality which makes their claim ambiguous. On the one hand, they clearly acknowledge the intensional meaning of conditional statements; similarly, they take a serious view on the effect of context and content on the interpretation of conditionals. All this is consistent with their claim. But on the other hand, starting with their definition of the core meaning in terms of possibilities which makes it identical with material implication, they carry on to define the various interpretations in terms of sets of mental models that are subsets of the models of the appropriate core meaning. The semantic and pragmatic modulations work like a filter which constrains the truth-functional meaning. Here lies the ambiguity for, after modulation, the various interpretations remain defined extensionally. In brief, the authors are right that their conditionals are not (purely) truth-functional, but the

semantics which they adopt is nevertheless extensional. This has the consequence that the probability of conditionals has to be that of the material conditional. Whether this is the probability that reasoners represent rather than the conditional probability of C given A is a question of major interest to which we will turn shortly.

2. 3. The probabilistic approach.

Oaksford and Chater (1991; Chater and Oaksford, 2000) argue that logic is inadequate to account for performance in reasoning tasks because reasoners use their everyday uncertain reasoning strategies, whose nature is probabilistic. This approach applies in a straightforward manner to the conditional reasoning task (Oaksford, Chater and Larkin, 2000). Reasoners are assumed to endorse the conclusion of a conditional argument in direct proportion to the conditional probability of the conclusion given the minor premise. With the following notations: $P(A) = a$, $P(C) = c$, and $P(\text{not-}C/A) = \varepsilon$, it follows that the probability associated with the major (conditional) premise is $1-\varepsilon$ (ε is called the *exception parameter),* which also provides the probability of endorsement of MP. The probabilities of endorsement of the other arguments are:

DA: $P(\text{not-}C/\text{not-}A) = 1-c-a\varepsilon / 1-a$

AC: $P(A/C) = a(1- \varepsilon) / c$

MT: $P(\text{not-}A/\text{not-}C) = 1-c-a\varepsilon / 1-c$

To test the model, the authors ran three experiments (two based on frequency distributions of artificial materials and one with everyday life categories and sentences). The aim was to test hypotheses related to negative bias (which is beyond the scope of his review). High rates of endorsement were predicted (and confirmed) with higher probability of the conclusion or, in some predictable cases, with lower probability of the minor premise. Similar predictions that were confirmed also apply to another set of four arguments with negated consequents.

Although these results support the model, the test is weak and fairly indirect. A strong test should consist of measuring or manipulating the parameters for a set of arguments in order to predict the rates of endorsement, and then comparing these with participants' ratings. Methodologically, the first two experiments exploit a very complicated scenario. All three experiments are self-serving in that the task is presented as a probabilistic task. As far as the model is concerned, there is a problem with the

exception parameter ε: once a realistic, small enough, value for it has been entered in the equation for MT, the rate of endorsement is much higher than usually observed. In order to obtain usual rates for MT, the value of ε should be increased, but now it is MP that has a rate of endorsement that is too low.

Oberauer**,** Weidenfeld and Hörnig (2004) tested the model in two experiments in which participants had to learn prior probabilities. The task followed closely the usual procedure and apart from a small effect for AC in one experiment, none of the predictions based on the magnitude of the prior probabilities of the conclusion was confirmed.

Finally, there is a serious shortcoming in the model. As the authors are aware, from a psychological point of view, a description at the algorithmic level is missing; this means that even assuming that the model is computationally correct, there is no hypothesis about the processes by which reasoners perform the computation.

In brief, although the model is appealing for its simplicity, in its present state it suffers from too many defects to be seriously considered. For further criticism and appraisal, see Schroyens and Schaeken (2003) and Oaksford and Chater's (2003a) reply. But in spite of its shortcomings, the probabilistic approach to human deductive reasoning has merit in pinpointing a variable of utmost importance, namely the uncertainty that affects information in daily life.

2. 4. The probability of conditional statements

One way of testing between the semantic approaches to conditionals is to consider how they define the probability of a conditional. Although the mental logic theory does not make an explicit claim, it seems to be a straightforward consequence of this theory that the probability of the conditional *if A, C* should be P(C/A), the conditional probability of C on A. This is because, as we noted earlier, the schema of conditional proof is a procedure that concerns itself with whether, and to which extent, C follows under the supposition that A, which is entirely compatible, if not identical with, the Ramsey test. For the mental model theory, the probability of a conditional statement should equal the probability of the material conditional P(not-A or B) for reasoners who construct full, explicit models, but a different value for those who construct an implicit model. Finally, the conditional probability is the obvious solution

for the probabilistic approach. We summarise the results of a few recent studies that used two kinds of materials, conventional or natural.

Studies of the first kind typically use fictitious cards bearing one of two figures (C, ¬C) that can be in two colours (A, ¬A), creating four cases whose frequencies can be manipulated so as to vary P(AC) and P(C/A). Participants have to evaluate the probability of the truth of statements such as *if the card is yellow then it has a circle.* Two studies run independently (Evans, Handley and Over, 2003; Oberauer and Wilmhelm, 2003) showed that across participants the main determinant of the evaluation was the statement's conditional probability, as indicated by the increase in *true* ratings with P(AC) and its decrease with P(A¬C). The probability of the associated material conditional could be unambiguously discarded as a determinant of the evaluation. Individual analyses revealed that one half of the participants had a conditional probability pattern of response whereas most others had an unexpected conjunctive pattern.

To avoid using objective probabilities inferred from frequency distributions, Over, Hadjichristidis, Evans, Handley and Sloman (in press) used more natural materials consisting of statements conveying predictions in economic or social domains, e.g., *if fertility treatment improves then the world population will increase.* Participants were asked an estimate of the probability of the antecedent, of the consequent, and of the truth of the statement. There was no evidence of conjunctive probabilities or of material conditional interpretations and very strong support for the conditional probability interpretation.

In brief, when reasoners are asked to estimate the probability of the truth of a conditional statement *if A then C* referring to frequency distributions that invites an extensional interpretation, their estimates are compatible with the conditional probability P(C/A) for about one half of them and with the conjunctive probability P(AC) for the other half. However, the latter conjunctive interpretation disappears with thematic statements that invite an intensional interpretation.

A defence of the mental model point of view is presented by Girotto and Johnson-Laird (2004) but their interpretation of the results seems highly debatable.

We conclude that it seems firmly established that when reasoners are asked the probability that a conditional statement is true the answers are first P(C/A), and second

P(AC) in the extensional case, but only the former in the intensional case. The conditional probability hypothesis seems strongly supported, but there remain a few interrogations related to the participants' understanding of the question about the truth of the conditional statement. This is because, on the one hand, supporters of the view that conditionals with false antecedents lack a truth value (e.g., Adams, 1975) may claim that the question is meaningless and that, consequently, it is re-interpreted in various ways, one of which would give rise to a conjunctive answer. On the other hand, there is a possibility that, as claimed by Girotto and Johnson-Laird (2004), the question is re-interpreted as *if A what is the probability that C* because people expect the question to bear on the main clause rather than the subordinate. They report, in a think aloud replication, clear cases of reformulation in terms of conditional probability. But of course this can be taken as evidence in favour of the conditional probability hypothesis, that is, for these participants the question does not make sense otherwise. Clearly, more experiments are needed to investigate these possibilities.

3. Reasoning with conditionals in rich contexts.

In this section, we consider research that has been developed in the last two decades on reasoning with conditional statements whose comprehension requires the exploitation of the reasoners' knowledge base to a much more considerable extent than in the impoverished contexts used to study formal reasoning. This research is closer to everyday reasoning where premises tend to be uncertain and inferences defeasible. We review the experimental situations --most of which use the conditional reasoning task-- and the main results, and then finish with an examination of how the main theories can explain these observations.

3.1. The deduction case: the effect of uncertain premises

From experience, reasoners are aware that their main sources of information, namely, verbal communication, perception, recall from memory, and inference, are fallible. They generally do not have full confidence in the premises of arguments. Moreover, even if their channels of information were perfectly reliable, they would still have to accept that information stored in their knowledge base is often intrinsically

characterised by a lack of certainty for several reasons: this information may be probabilistic in nature, or it may be imprecise in nature, or it may concern generalisations that have exceptions. In brief, in real life arguments typically start with uncertain premises, and this applies in particular to deductive arguments such as MP and MT.

3.1.1. The propagation of uncertainty

That the uncertainty of premises affects the conclusion of conditional reasoning arguments can be shown most simply either by qualifying the consequent of the conditional by a probability term, in which case on observes that reasoners qualify the conclusion (generally by the same term: George, 1997), or by asking participants to rate their confidence in the premise, and then in the conclusion. In one of George's (1995) experiments, one half declared the conclusion true, whereas the other half seldom or never did so, which suggests that, at least in an experimental setting, people function in either one of two modes: a trustful mode in which they assume the truth of the premises, even though these are controversial, before they engage in a standard deduction; and a distrustful mode in which they accept the uncertainty of the premises, which results in an uncertain conclusion. Among the latter participants, the degrees of confidence in the premise and in the conclusion were highly correlated, which suggests an increasing function. This was confirmed by studies exploiting common knowledge to vary the credibility of the premises.

Thompson (1994, 1995) considered conditional premises expressing various relations, such as causes, obligations, permissions and definitions. The rate of endorsement of the conclusion was found to be an increasing function of the conditional probability of the consequent on the antecedent (independently estimated by judges). Similar results were obtained by Liu, Lo, and Wu (1996) with conditionals expressing definitions, regulations or stereotypes. Cummins (1995; Cummins, Lubart, Alksnis, and Rist, 1991) studied MP and MT arguments whose major premise was a causal rule. She compared arguments that differed by the number of their disabling conditions, that is, factors that can prevent an effect from occurring. Assuming that the confidence in the conditional decreases as this number increases, it can be expected that the confidence in the conclusion will be a decreasing function of the number of disabling conditions, which

indeed was observed. For example, people accept less readily the conclusion of MP with *if the match was struck, then it lit*, which has many disabling conditions than with *if Joe cut his finger,* which has few.

In summary, there is a propagation of the level of uncertainty of the conditional premise of MP and MT arguments to the conclusion, and this follows an increasing function.

3. 1. 2. Defeating a conclusion

Whereas in the previous situations the uncertainty is manipulated by exploiting conditionals that have some given credibility, another situation has been investigated which corresponds to "defeasible reasoning": the uncertainty is manipulated by introducing an additional premise, which results in the suspension or the revocation of a conclusion that could have initially been drawn. In all these manipulations the additional premise refers to a hitherto unstated condition which turns out to be necessary for the consequent to hold, a condition which we will call a "complementary necessary condition" (in short a CNC), generalising the notion of enabling or disabling condition.

3.1.2.1. The suppression effect.

We begin with the so-called "suppression effect" which has been much debated in the psychological literature and beyond, and which further analysis has shown to be a particular --and more complicated case -- of a more general phenomenon.

Byrne (1989) presented one control group with standard MP (and MT), e. g., *if she has an essay to write then she will study late in the library; she has an essay to write;* therefore: (a) *she will study late in the library;* (b) *she will not study late in the library;* (c) *she may or may not study late in the library.* Another group received the same arguments modified by an additional conditional premise that had the same consequent as the major premise and an antecedent that was a CNC of the major premise, like *if the library stays open then she will study late in the library.* While most participants in the first group endorsed the conclusion, for the second group, there was an apparent "suppression" of the conclusions, as the endorsement rate collapsed to around 35 percent for both MP and MT. The results have been replicated many times with endorsement rates closer to 50% on the average. Interestingly, this experiment was not presented by its author as an investigation of reasoning under uncertainty, and indeed the task does not have the specific features of the tasks reviewed so far: the

major conditional premise is not modified by a probability or a frequency term, nor is its credibility intrinsically questionable in the context that is provided.

Politzer and Braine (1991) proposed that the decrease in the rate of endorsement of the conclusion could be regarded as an instance of deduction under uncertainty and was explainable within a pragmatic framework. They remarked that the additional premise does not state categorically that a necessary condition is unfulfilled, but rather is a conditional which mentions the condition and conversationally implicates (Grice, 1989; Levinson, 2000) that it might not be fullfilled. Whereas in the standard argument the major conditional premise is uttered with a *ceteris paribus* assumption of normality (the CNC must be satisfied if one is to believe the premise), in the modified argument this assumption is questioned by the epistemic implicature just mentioned, and the major premise is no longer certain: consequently, the conclusion is no longer warranted. This "pragmatic epistemic explanation" of the suppression task has received considerable support.

Politzer (2005) observed, in a production task, that when the implicit expression of uncertainty about the satisfaction of a CNC conveyed by the additional premise *(if the library stays open then she will study late)* is replaced by an explicit expression *(it is not certain that the library stays open),* the response patterns are virtually identical, confirming that this uncertainty is at the root of the suppression effect. Stevenson and Over (1995) observed that the rate of endorsement of the conclusion, as well as the certainty ratings, increased with the likelihood of the satisfaction of the CNC. Similar results were obtained by Manktelow and Fairley (2000) who manipulated the degree of satisfaction of the CNC.

Politzer and Bourmaud (2002) generalised these results by using a variety of conditionals (means-ends, causal, remedial, decision statements) with MT arguments and by combining the manipulation of the degrees of satisfaction of the CNC with the degrees of necessity of the CNC (some are absolutely necessary, others are not). There was a strong and reliable correlation between credibility levels of the conclusion and credibility of the major premise. This verified again and generalised the notion that the confidence in the conclusion of simple deductive arguments is an increasing function of the degree to which a CNC is satisfied and of its level of necessity, which can be

understood if the CNC acts as a mediator for the credibility of the major conditional premise. This latter claim is defended in Politzer (2005).

In brief, the task exhibiting the "suppression effect" can be subsumed under the other, more simple, tasks mentioned in this section, in which a doubt is introduced about the credibility of the major conditional premise, not by use of another conditional, but by questioning the satisfaction of a CNC either directly or indirectly.

3.1.2.2. Consequential conditionals

Another case of revocation of the conclusion of MP has been described by Bonnefon and Hilton (2004) with "consequential" conditionals, which are such that the antecedent is an action, and the consequent a desirable or an undesirable outcome of this action. The authors have shown that their utterance pragmatically invites the inference to the truth or falsity of their antecedent. For instance, given *if Mary has herTV fixed, she will not be able to pay the electricity bill,* the majority of people expect *Mary will not have her TV fixed.* Now, take the premises *if Mary's TV is broken, she will have it fixed; Mary's TV is broken:* most people conclude by MP, *Mary will have her TV fixed.*. But when reasoners were presented with the additional consequential premise *if Mary has her TV fixed, she will not be able to pay the electricity bill,* which invites an inference to the negated conclusion (*Mary will not have her TV fixed)*, the rate of endorsement of the conclusion collapsed by one half. Moreover, the more undesirable the outcome, the larger the effect.

3.1.2.3. Inferential conditionals

We consider a last case of defeasible reasoning in the framework of MP or MT arguments, that is based on the existence of an alternative to a justification previously formulated by a conditional sentence. Pollock (1987) identified two reasons (called "defeaters") that may lead to the revocation of a conclusion. Generally speaking, given a reason P to believe Q, a "defeater" for P is a reason R which is logically consistent with P but such that (P&R) is not a reason to believe Q. A "rebutting defeater" R for P is a defeater which is a reason for believing not-Q; and an "undercutting defeater" R for P is a defeater which is a reason for denying that P would not be true unless Q were true.

In terms of this conceptualisation, the effects considered in the last two sections can be regarded as cases of the operation of rebutting defeaters. What about undercutting defeaters? Consider the kind of conditionals called "inferential" (Sweetser,

1990, Dancygier, 1998) which express the speaker's belief in the consequent, given that its truth would be a good explanation for the antecedent. Inferential conditionals often contain causal relations, such as *if Peter is late, he was caught in a traffic jam.* (In terms of causality, inferentials and their direct counterparts are equivalent to Pearl's (1991) "evidentials" and "causals", respectively). Given the inferential premise *if Mary is in the library, she has an essay to write,* an undercutting defeater is provided by *if Mary wants to borrow a book, she is in the library*.

Now take the premises. *if Mary is in the library, she has an essay to write; Mary is in the library:* most people are highly confident in the conclusion *Mary has an essay to write*. But when reasoners were presented with the additional premise *if Mary wants to borrow a book, she is in the library* (or *Mary might want to borrow a book)*, the level of confidence in the conclusion decreased significantly (Politzer and Bonnefon, 2006), showing that undercutters have the potential to revoke the conclusion of MP (and also MT) arguments whose major premise is an inferential.

3.1.2.4. Theoretical accounts.

Formally, the foregoing three kinds of manipulation share an important characteristic. Let *if A, C* be the major premise of the argument; let the additional premises of the standard revocation task and of the "inferential conditional" task be *if N, C,* and *if N, A,* respectively, and finally let the additional consequential premise be *if C, N* in which N is a necessary condition, an alternative explanation, or an undesirable consequence, respectively, with respect to the occurrence of C. Then, rewriting the major premise as *if (A & N), C,* in the former case, or *if (A & not-N), C,* in the latter two, it appears that questioning the satisfaction, or the non-satisfaction of N is the source of the uncertainty of the premise in the three cases. (That is, with the same examples as above, the premises are rewritten, respectively, as: (i) *if Mary has an essay to write and the library is open (then) she will study in the library;* (ii) *if Mary studies in the library and she does not want to borrow a book (then) she has an essay to write;* (iii) *if Mary's TV is broken and she will be able to pay the electricity bill (then) she will have it fixed).* In brief, all three indirect manipulations of uncertainty are amenable to the same formal representation of the major premise of the argument; this premise incorporates in its antecedent the source of the uncertainty which initially was "idle" in the encyclopedic knowledge base and then activated by the additional premise.

In rewriting the major premise *if A, C,* we have made apparent the common form of *strengthening of the antecedent, if A&B, C.* In sum, the uncertain situations that lead to various degrees of doubt about the conclusion are qualitative variants of the standard sure case where the conclusion is accepted. This may cover various degrees of doubt in the conclusion, up to complete uncertainty. Reasoners are sensitive to this continuity, and theories of reasoning must account for both the uncertain and the certain judgments.

However, this has a symmetrical counterpart in the paradoxical situation of strengthening usually considered to exemplify nonmonotonicity: the conjunct *B* is replaced by its negation (that is, the CNC is not satisfied) and the conditional *if A&not-B, C* is not assertable; or, as a variant, taking the simple conditional *if A, C* and adding *not-B* to the minor premise *A,* the conclusion of the MP has to be denied.

The account by the Mental Models theory of this latter case of stengthening of the antecedent has been given in section 2.2.2 (applied to MP). It faces a first difficulty: it assumes that precedence is given to general knowledge over the fact at hand, so that [match-soaked ¬match-lights] takes precedence over [match-struck match-lights]. But it is unclear why the latter should not belong to general knowledge; clearly another principle should operate, or the solution comes on a case by case basis dictated by the knowledge base.

More seriously, when applied to MT, the explanation predicts a final model that yields: *the match is soaked in water and does not light and is not struck,* and the conclusion *the match is not struck.* However, about two thirds of the reasoners draw the conclusion that one cannot know whether or not the match was struck. Maybe the relevant fact in the knowledge base is that if a match is not struck it does not light, but it can be shown that this yields exactly the same conclusion. The mental models theory cannot account for conditional strenghtening, or for the associated phenomena such as the suppression effect. In previous versions of the theory, Byrne, Espino and Santamaría (1999) attempted to account for the suppression effect in terms of counterexample; but this explanation made entirely wrong predictions, as demonstrated in Politzer (2005).

In mental logic, a conditional cannot be strengthened by application of the constraint on the CP so that *if A&not-B, C* is blocked. But the theory faces the same difficulty as the mental models theory to predict the revocation of the conclusion for

MP. The conclusions intuitively acceptable *may be C* or *not-C* depend on the knowledge base and the theoretical approaches developed for formal reasoning meet their limitation. For instance, in the well-known case of the "Nixon diamond" (if one is a Quaker, one is a pacifist; if one is a Republican, one is not a pacifist; Nixon is a Quaker and a Republican), the conclusion that Nixon is a pacifist, may be a pacifist or is not a pacifist depends on which of the premises takes precedence over the other.

Stenning and van Lambalgen (2005) presented a nonmonotonic model using logic programming with negation as failure. This work contains both a simple formalisation and a proposal for an implementation in terms of connexionist network. One key assumption of the model is the representation of the main premise as *if A & ¬ab, C* where *ab* is a proposition which indicates that something abnormal is the case. The authors show that their approach predicts the right results for MP and MT, that is, no conclusion follows in the absence of information about the CNC, which is represented in this formalism by *if not N, ab.* This nicely captures the phenomenon but misses the qualitative results linked with degrees of doubt in the conclusion.

On the other hand, Oaksford and Chater (2003b) applied their conditional reasoning equations to the revocation tasks (Cummins's and Byrne's). The key idea is that bringing to mind additional necessary factors (that is, CNC's) increases the exception parameter. In spite of the shortcomings of the model, the interest of this kind of approach is that it takes a serious view on reasoners' graded evaluations of conclusions, which non probabilistic approaches cannot do.

It will be apparent that the psychological issues discussed in this section are highly related to work and issues encountered by AI investigators of reasoning with imprecise or uncertain information. We mention two studies that aim to test the psychological plausibility of the set of rationality postulates for nonmonotonic reasoning known as "system P" (Kraus, Lehmann and Magidor, 1990) whose results, by and large, support the hypothesis: Da Silva Neves, Bonnefon and Raufaste (2002), and Pfeifer and Kleiter (2005a).

3. 2. The invalid arguments

We now consider performance on the invalid arguments AC and DA. The manipulation of Rumain, Connell, and Braine (1983) described in section 2.1.2 was

specially designed to alert the participants to the plurality of antecedents so that the invited inferences be countermanded. The subsequent investigations with thematic materials, a few of which will be described, have followed the same principle.

In Cummins' studies of causal conditionals described earlier, the role of alternative causes on invalid arguments was also investigated. For example, comparing two AC arguments such as *If the match was struck, then it lit; the match lit;* therefore *the match was struck* and *If Mary jumped into the swimming pool, then she got wet; Mary got wet* therefore *she jumped into the swimming pool,* people were more prone to accept the conclusion of the first, whose major premise has few alternative causes, than that of the second, whose major premise has many. Thompson (1994, 1995) classified conditionals on the basis of their "level of necessity", that is, the extent to which the consequent occurs only when the antecedent occurs. She observed that the rate of endorsement of AC and DA arguments was an increasing function of this level of necessity. These studies show that the more available the alternative antecedents, (low number of alternative causes, high level of necessity) the less likely the invited inference that leads to the endorsement of the conclusion.

Quinn and Markovits (1998) compared causal conditionals that differed in the strength of the association between antecedent and consequent defined as follows. Given an effect, judges were requested to produce as many causes as they could in a limited time. Considering two causes produced for an effect, the more frequent was considered as the more strongly associated to the effect, so that the authors could define two groups of conditionals, a strong group (e. g. *If a dog has fleas, then it will scratch constantly)* and a weak group (e. g. *If a dog has a skin disease, then it will scratch constantly).* There were fewer endorsements of the conclusion of AC and DA arguments with the latter than with the former. With the weak association, the antecedent is not the most available; therefore it is relatively easy for a more available antecedent to be retrieved and play the role of an alternative cause. In contrast, with the strong association, the antecedent is the most available; it is therefore relatively difficult for a less available antecedent to be retrieved and play the role of an alternative cause.

De Neys, Schaeken and d'Ydewalle (2003) presented participants with conditional statements and asked them to produce as many alternative antecedents as they could. One month later they were presented with the corresponding AC and DA arguments

and asked to rate the conclusion (on a 7-point certainty scale from *cannot be drawn* to *can be drawn).* There was a stepwise decrease in certainty as a function of the number of alternatives produced in the initial phase.

In sum, reasoners are very sensitive to alternative antecedents (or, more generally, to alternative justifications, as shown earlier for inferentials in which the alternative refers to the consequent). One finds, however, across experiments, about a quarter of participants who apparently maintain the fallacious answers even when the alternatives are available from long term memory. Are these answers genuinely fallacious? This seems doubtful. For one thing, participants in the experiments are likely to assume that, because the experimenter has selected one antecedent out of all the others, this particular antecedent is relevant and must be used in the answer. More importantly, the AC argument coincides with an abduction (in its elementary, restricted, sense). As there has been increasing interest in abduction in AI (Josephson and Josephson, 1994), in philosophy of science (Magnani, 2001) as well as in argumentation theory (Walton, 2004), it might be time for psychologists to adopt this framework to study the fallacies. Now, in view of the pervasiveness of abduction in daily life, if the instructions to draw a conclusion that follows logically are understood as drawing a plausible conclusion, participants are justified in giving the antecedent as a response. These remarks also apply to the formal task, in which an additional layer of deception is added in that the set of alternatives is not clearly defined, except for not-A. In brief, it might very well be the case that the endorsement of AC and DA is the experimenters' fallacy rather than the reasoners'.

Conclusion

Conditional reasoning highlights the challenge that theories of reasoning are facing from the moment they aim to explain both formal and informal reasoning. The two main theoretical approaches to deductive reasoning have been developed to account for formal reasoning and it is not clear whether they could be extended to deal with uncertain reasoning and defeasible reasoning. This is particularly clear for the mental models theory which makes arbitrary or wrong predictions even for some of the most simple arguments. Both theories need to define how to represent the probabilities attached to

propositions and mental logic would also need a way to allow for potential abnormal conditions to be represented; this might be feasible (Politzer and Bourmaud, 2002) but no project of this kind has been realised. The probabilistic approach has the opposite problem; it can in principle deal with informal reasoning (although no proposal that has psychological plausibility has been developed) but does not account for formal competence such as deriving simple proofs in propositional reasoning. Reasoners perform logic-based inferences but at the same time they attribute degrees of belief to their conclusions; their assessments are often flexible and fined-grained. For example, in a paradigm borrowed from AI, namely belief revision (Gärdenfors, 1988), which space does not allow us to cover, when reasoners decide to give up a premise, they do so 80% of the time by expressing doubt about the premise rather than by denying it (Politzer and Carles, 2001). In sum, standard logic-based formalisms are unable to accommodate people's degrees of belief, and standard probability-based formalisms are not appropriate to account for people's elementary deductive competence. As the theories of deduction on which this review has focused have been developed for quite some time, the views of logic or probability to which they refer are inevitably fairly standard. To develop new theoretical approaches, psychologists could borrow conceptual tools from formalisms such as nonmonotonic logics, possibilistic logic, fuzzy logic, multi-valued logics and the like (see Smets, 1998). Another promising direction is offered by probability logics (Adams, 1998; Bacchus, 1990; Hailperin, 1996) which can deal with both facets of human deductive competence, viz. the derivation of proofs and the management of uncertainty. (For the outline of a specific application, see Politzer, 2005, and for a theoretical approach, see Pfeifer and Kleiter, 2005b, in press).

CNRS, Institut Jean-Nicod, Paris

References

Adams, E. : 1975, *The logic of conditionals,* Dordrecht: Reidel.

Adams, E. : 1998, *A primer of probability logic.* Stanford, CSLI Publications.

Bacchus, F.: 1990, *Representing and reasoning with probabilistic knowledge,* Cambridge, MIT Press.

Bell, V. A., Johnson-Laird, P. N.: 1998, 'A model theory of modal reasoning', *Cognitive Science* **22,** 25-51.

Bonnefon, J.-F., Hilton, D. J.: 2004, 'Consequential conditionals: Invited and suppressed inferences from valued outcomes', *Journal of Experimental Psychology: Learning, Memory and Cognition* **30,** 28-37.

Braine, M. D. S., O'Brien, D. P.: 1991, 'A theory of IF: A lexical entry, reasoning program, and pragmatic principles', *Psychological Review* **98,** 182-203.

Braine, M. D. S., O'Brien, D. P.: 1998, *Mental logic,* Mahwah, N. J. : Lawrence Erlbaum.

Braine, M. D. S., Reiser, B. J. , Rumain, B.: 1984, 'Some empirical justification for a theory of natural propositional logic', in G. H. Bower (ed.), *The psychology of learning and motivation,* Vol. 18. N. Y. : Academic Press, pp. 313-371.

Byrne, R. M. J.: 1989, 'Suppressing valid inferences with conditionals', *Cognition* **31,** 61-83.

Byrne, R. M. J. , Espino, O. , & Santamaría, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language, 40,* 347-373.

Byrne, R. M. J. & Johnson-Laird, P. N.: 1989, 'Spatial reasoning', *Journal of Memory and Language* ***28,*** 564-575.

Chater, N., Oaksford, M.: 2000, 'The rational analysis of mind and behavior', *Synthese* **122,** 93-131.

Cummins, D. D.: 1995, 'Naive theories and causal deduction', Memory and *Cognition* **23,** 646-658.

Cummins, D. D., Lubart, T., Alksnis, O., Rist, R.: 1991, 'Conditional reasoning and causation', *Memory and Cognition* **19,** 274-282.

Dancygier, B.: 1998, *Conditionals and prediction. Time, knowledge, and causation in conditional constructions,* Cambridge: Cambridge University Press.

Da Silva Neves, R., Bonnefon, J.-F., Raufaste, E.: 2002, 'An empirical test of patterns for nonmonotonic inference', *Annals of Mathematics and Arificial Intelligence* **34,** 107-130.

Delval, J. A., Riviére, A.: 1975, ' "Si llueve, Elisa lleva sombrero": Una investigacion psicologica sobre la tabla de verdad del condicional', *Revista de Psicologia General y Aplicada* **30,** *136,* 825-850.

De Neys, W., Schaeken, W., d'Ydewalle, G.: 2003, 'Inference suppression and semantic memory retrieval: Every counterexample counts', *Memory and Cognition* **31,** 581-595.

Evans, J. St. B. T.: 1972, 'Reasoning with negatives', *British Journal of Psycholog y* **63,** 213-219.

Evans, J. St. B. T.: 1996, Deciding before you think: Relevance and reasoning in the selection task' , *British Journal of Psychology* **87,** 223-240.

Evans, J. St. B. T., Legrenzi, P., Girotto, V.: 1999, 'The influence of linguistic form on reasoning: The case of matching bias', *Quarterly Journal of Experimental Psychology* **52A,** 185-216.

Evans, J. St. B. T., Handley, S. J., Over, D. E.: 2003, 'Conditionals and condtional probability', *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **29,** 321-335.

Evans, J. St. B. T., Over, D. E.: 2004, '*If',* Oxford: Oxford University Press.

Fillenbaum, S.: 1975, 'A note on memory for sense: Incidental recognition of warning phrased as conditionals, disjunctives, and conjunctives', *Bulletin of the Psychonomic Society* **6,** 293-294.

Fillenbaum, S.: 1976, 'Inducements: On the phrasing and logic of conditional promises, threats, and warnings', *Psychological Research* **38,** 231-250.

Gärdenfors, P.: 1988, *Knowledge in flux,* Cambridge, MIT Press.

Gazdar, G.: 1979). *Pragmatics. Implicature, presupposition and logical form,* New York : Academic Press.

Geis, M. L., Zwicky, A. M.: 1971, 'On invited inferences', *Linguistic Inquiry* **2,** 561-566.

George, C.: 1995, 'The endorsement of the premises: Assumption-based or belief-based reasoning', *British Journal of Psychology* **86,** 93-111.

George, C.: 1997, 'Reasoning from uncertain premises', *Thinking and Reasoning* **3,** 161-189.

Girotto, V., Johnson-Laird, P. N.: 2004, 'The probability of condirtionals', *Psychologia* **47,** 207-225.

Grice, H. P.: 1975, 'Logic and conversation', in P. Cole, J. L. Morgan (eds.), *Syntax and Semantics.* Vol 3: Speech acts, New York : Academic Press.

Grice, P.: 1989, *Studies in the way of words,* Cambridge: Harvard University Press.

Hailperin, T.: 1996, *Sentential probability logic,* Bethlehem, Lehigh University Press.

Horn, L. R.: 1989, *A natural history of negation,* Chicago: University of Chicago Press.

Johnson-Laird, P. N.: 1999, 'Deductive reasoning', *Annual Review of Psychology* **50,** 109-135.

Johnson-Laird, P. N.: 2001, 'Mental models and deduction', *Trends in Cognitive Sciences* **5,** 434-442.

Johnson-Laird, P. N., Bara, B. G.: 1984, 'Syllogistic inference', *Cognition* **16,** 1-61.

Johnson-Laird, P. N., Byrne, R. M. J.: 1991, *Deduction,* Hove: Lawrence Erlbaum.

Johnson-Laird, P. N., Byrne, R. M. J.: 2002, 'Conditionals: a theory of meaning, pragmatics, and inference' , *Psychological Review* **109,** 646-678.

Johnson-Laird, P. N. Tagart, J.: 1969, 'How implication is understood', *American Journal of Psychology* **82,** 367-373.

Josephson, J. R. , & Josephson, S. G.: 1994, *Abductive inference. Computation, philosophy, technology,* Cambridge, C. U. P.

Kraus, S., Lehmann, D., Magidor, M.: 1990, 'Nonmonotonic reasoning, preferential models and cumulative logics' , *Artificial Intelligence* **44,** 167-207.

Levinson, S. C. (2000 , *Presumptive meanings,* Cambridge: MIT Press.

Liu, I.-M., Lo, K.-C., Wu, J.-T.: 1996, 'A probabilistic interpretation of "If-Then"', *Quarterly Journal of Experimental Psychology* **49A,** 828-844.

Magnani, L.:2001, *Abduction, reason, and science: Processes of discovery and explanation,* New York, Kluwer.

Manktelow, K., Fairley, N.: 2000, 'Superordinate principles in reasoning with causal and deontic conditionals', *Thinking and Reasoning* **6,** 41-65.

Noveck, I. A., Sperber, D.: 2004, *Experimental pragmatics,* Houndmills: Palgrave McMillan.

Oaksford, M. Chater, N.: 1991, 'Against logicist cognitive science', *Mind and Language* **6,** 1-38.

Oaksford, M., Chater, N.: 1994, 'A rational analysis of the selection task as optimal data selection', *Psychological Review* **101,** 608-631.

Oaksford, M., Chater, N.: 2003a, 'Computational levels and conditional inference: Reply to Schroyens and Schaeken' , *Journal of Experimental Psychology: Learning, Memory, and Cognition* **29,** 150-156.

Oaksford, M., Chater, N., 2003b: 'Probabilities and pragmatics in conditional inference: Suppression and order effects, in D. Hardman, L. Macchi (eds.), *Thinking: Psychological perspectives on reasoning, judgment and decision making*, Chichester: John Wiley, pp. 95-122

Oaksford, M., Chater, N., Larkin, J.: 2000, 'Probabilities and polarity biases in conditional inference', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **26,** 883-899.

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

Oberauer, K., Weidenfeld, A., Hörnig, R.: 2004, 'Logical reasoning and probabilities: A comprehensive test of Oaksford and Chater (2001)', *Psychonomic Bulletin and Review* **11,** 521-527.

Oberauer, K. , Wilhelm, O.: 2003, 'The meaning(s) of conditionals: Conditional probabilities, mental models, and personal utilities', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **29,** 680-693.

Over, D. E., Hadjichristidis, C., Evans, J. St. B. T., Handley, S. J., Sloman, S. A. (in press): 'The probability of ordinary indicative conditionals', *Cognitive Psychology.*

Paris, S. G.: 1973, Comprehension of language connectives and propositional logical relationships', *Journal of Experimental Child Psychology* **16,** 278-291.

Pearl, J.: 1991, *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* 2nd ed. San Francisco: Morgan Kaufmann.

Pfeifer, N., Kleiter, G. D., (in press), 'Inference in conditional probability logic', *Kybernetica*.

Pfeifer, N., Kleiter, G. D.: 2005a 'Coherence and nonmonotonocity in human reasoning', Synthese **146,** 93-109.

Pfeifer, N. , Kleiter, G. D.: 2005b. 'Towards a mental probability logic', *Psychologica Belgica* **45,** 71-99.

Poletiek, F. H.: 2001, *Hypothesis-testing behaviour.* Hove: Psychology Press.

Politzer, G.: 1981, 'Differences in interpretation of implication', *American Journal of Psycholog y* **94,** 461-477.

Politzer, G.: 2004, 'Reasoning, judgement and pragmatics' , in I. N. Noveck, D. Sperber (eds.), *Experimental pragmatics*, Houndmills: Palgrave, pp.94-115

Politzer, G.: 2005, 'Uncertainty and the suppression of inferences', *Thinking and Reasoning* **11,** 5-33.

Politzer, G. Bonnefon, J.-F.: 2006, 'Two varieties of conditionals and two types of defeaters help reveal two fundamental types of reasoning', *Mind and Language* **21,** 484-503.

Politzer, G., Bourmaud, G.: 2002, 'Deductive reasoning from uncertain conditionals', *British Journal of Psycholog y* **93,** 345-381.

Politzer, G., Braine, M. D. S.: 1991, 'Responses to inconsistent premisses cannot count as suppression of valid inferences', *Cognition* **38,** 103-108.

Politzer, G., Carles, L.: 2001, 'Belief revision and uncertain reasoning', *Thinking and Reasoning* **7,** 217-234.

Politzer, G., Macchi, L.: 2000, 'Reasoning and pragmatics', *Mind and Society* **1,** 73-93.

Pollock, J. L.: 1987, 'Defeasable reasoning' , *Cognitive Science* **11,** 481-518.

Quinn, S., Markovits, H.: 1998, 'Conditional reasoning, causality, and the structure of semantic memory: strength of association as a predictive factor for content effects', *Cognition* **68,** B93-B101.

Ramsey, F. P.:1931, 'General propositions and causality', in R. B. Braithwaite (ed.), *The foundations of mathematics and other logical essays.* London: Routledge & Kegan Paul, pp. 237-255.

Rips, L. J.: 1994, *The psychology of proof,* Cambridge, Ma: M.I.T. Press.

Rumain, B., Connell, J., Braine, M. D. S.: 1983, 'Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: *If* is not the biconditional', *Developmental Psycholog y* **19,** 471-481.

Scholnick, E. K., Wing, C. S.: 1991, 'Speaking deductively: Preschoolers' use of *If* in conversation and in conditional inference', *Developmental Psycholog y* **27,** 249-258.

Scholnick, E. K., Wing, C. S.: 1992, 'Speaking deductively: Using conversation to trace the origins of conditional thought in children', *Merrill-Palmer Quarterly* **38,** 1-20.

Schroyens, W., Schaeken, W. 2003: ' A critique of Oaksford, Chater, and Larkin's (2000) conditional probability model of conditional reasoning', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **29,** 140-149.

Smets, P.: 1998, (ed.), *Quantified representation of uncertainty and imprecision,* Vol. 1 of D. M. Gabbay & P. Smets (eds.), *Handbook of defeasible reasoning and uncertainty management systems*, Dordrecht, Kluwer.

Sperber, D., Cara, F., Girotto, V.: 1995, 'Relevance theory explains the selection task', *Cognition* **57,** 31-95.

Sperber, D., Wilson, D.: 1986, *Relevance: Communication and cognition,* London: Blackwell. [2nd ed., 1995].

Stalnaker, R. C., Thomason, R. H.: 1970, 'A semantic analysis of conditional logic', *Theoria* **36,** 23-42.

Stenning, K., Van Lambalgen, M.: 2004, 'A litle logic goes a long way: basing experiment on semantic theory in the cognitive science of conditional reasoning', *Cognitive Science* **28,** 481-529.

Stenning, K., Van Lambalgen, M.: 2005, 'Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task', *Cognitive Science* **29,** 919-960.

Stevenson, R. J., Over, D. E. : 1995, 'Deduction from uncertain premises', *Quarterly Journal of Experimental Psychology* **48A,** 613-643.

Sweetser, E.: 1990, *From etymology to pragmatics,* Cambridge, MA: Cambridge: Cambridge University Press.

Thomason, R. H.: 1970, 'A Fitch-style formulation of conditional logic', *Logique et Analyse* **52,** 397-412.

Thompson, V. A.: 1994, 'Interpretational factors in conditional reasoning', *Memory and Cognition* **22,** 742-758.

Thompson, V. A.: 1995, 'Conditional reasoning: The necessary and sufficient conditions', *Canadian Journal of Experimental Psychology* **49,** 1-60.

Walton, D.: 2004, *Abductive reasoning,* Tuscaloosa, University of Alabama Press.

Wason, P. C. : 1966, 'Reasoning'. In B. M. Foss (ed.), *New horizons in psychology.* Vol. 1, Pelican Book.