Reconciling probabilistic approaches and experimental pragmatics:  
The case of conditionals

Ira Noveck

Laboratoire sur le Langage le Cerveau et la Cognition (L2C2)  
Lyon, France
In the old days (circa 1980’s), the reasoning world was pretty straightforward
The question was about representation.
<table>
<thead>
<tr>
<th>Prémisses du schéma</th>
<th>Conclusion du schéma</th>
<th>Aussi connu comme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not not $P$</td>
<td>$P$</td>
<td>Double negation</td>
</tr>
<tr>
<td>If $P$ or $Q$ then $R$ ; $P$</td>
<td>$R$</td>
<td>Disjunctive Modus Ponens</td>
</tr>
<tr>
<td><strong>P or $Q$; not $P$</strong></td>
<td>$Q$</td>
<td><strong>Disjunction elimination</strong></td>
</tr>
<tr>
<td>Not both $P$ and $Q$ ; $P$</td>
<td>not-$Q$</td>
<td>Negated disjunction</td>
</tr>
<tr>
<td>$P$ or $Q$ ; If $P$ then $R$ ; If $Q$ then $R$</td>
<td>$R$</td>
<td></td>
</tr>
<tr>
<td>$P$ or $Q$ ; If $P$ then $R$ ; If $Q$ then $S$</td>
<td>$R$ or $S$</td>
<td></td>
</tr>
<tr>
<td><strong>If $P$ then $Q$ ; $P$</strong></td>
<td>$Q$</td>
<td><strong>Modus Ponens</strong></td>
</tr>
<tr>
<td>$P$ ; $Q$</td>
<td>$P$ and $Q$</td>
<td>Conjunction introduction</td>
</tr>
<tr>
<td>$P$ and $Q$</td>
<td>$P$</td>
<td>Conjunction elimination</td>
</tr>
<tr>
<td>$P$ and (Q or R)</td>
<td>(P and Q) or (P and R)</td>
<td></td>
</tr>
<tr>
<td>$P$ and not-$P$</td>
<td>Incompatibility</td>
<td>Contradiction</td>
</tr>
<tr>
<td>$P$ or $Q$ ; Not $P$ ; Not $Q$</td>
<td>Incompatibility</td>
<td>Contradiction</td>
</tr>
</tbody>
</table>
• These inferences are easy
• Applied in a programmatic order
• Provide people with inferences automatically.

I tell you that on the blackboard.

There is a C or T.
There is no T.
If there is a C then there is a G.
There is not both a G and a D.

And I ask you to evaluate as True or False:
There is a D.

Disjunction elimination
Modus Ponens
Negated conjunction
Contradiction

\[
\begin{align*}
\text{C or T; no T} & : \text{C} \\
\text{C \rightarrow G; C} & : \text{G} \\
\neg(G \& D; G) & : \neg D \\
D \& \neg D & : \text{Contradiction}
\end{align*}
\]
## Mental Models

<table>
<thead>
<tr>
<th>Connective</th>
<th>Initial</th>
<th>Explicit</th>
</tr>
</thead>
<tbody>
<tr>
<td>p and q</td>
<td>$p$</td>
<td>$q$</td>
</tr>
<tr>
<td></td>
<td>$p$</td>
<td>$q$</td>
</tr>
<tr>
<td>p or q</td>
<td>$p$</td>
<td>$q$</td>
</tr>
<tr>
<td>If $p$ then $q$</td>
<td>$[p]$</td>
<td>$q$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inclusive</th>
<th>Exclusive</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>$q$</td>
</tr>
<tr>
<td>$p$</td>
<td>not-$q$</td>
</tr>
<tr>
<td>not-$p$</td>
<td>$q$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional</th>
<th>Biconditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>$q$</td>
</tr>
<tr>
<td>not-$p$</td>
<td>not-$q$</td>
</tr>
<tr>
<td>not-$p$</td>
<td>$q$</td>
</tr>
</tbody>
</table>
There is a circle or a triangle  

There is no circle

1.

There is a circle or a triangle  

2.

3.

Conclusion:
Propositional Reasoning by Model?

Luca Bonatti

Two theories of propositional deductive reasoning are considered: Johnson-Laird’s mental models and Braine’s mental logic. The mental model theory, unlike Braine’s mental logic, offers a general theory of conditions under which the most important aspects of Braine’s theory are difficult to overcome. To understand them, I argue that (a) the model theory is biased in a way that it is difficult to overcome. (b) although these phenomena are not represented in the algorithm proposed to implement it (the use of invalid premises or any psychologically realistic premise), the effects of all these phenomena are not represented in the algorithm proposed to implement it, (c) Braine’s theory accounts for all of them, and (d) the model theory can predict Braine’s results only at the cost of even more simplification. I conclude that the mental model theory of propositional reasoning offers no reason to reject the program of mental logic.

Why Models Rather Than Rules Give a Better Account of Propositional Reasoning: A Reply to Bonatti and to O’Brien, Braine, and Yang

P. N. Johnson-Laird, Ruth M. J. Byrne, and Walter Schaeken

O’Brien, Braine, and Yang argue that the mental model theory of propositional reasoning is easy to refute, and they report 1 experiment that they believe falsifies the theory. In contrast, Bonatti argues that the model theory is too flexible to be falsified. We show that O’Brien et al.’s experiments do not refute the model theory and that Bonatti’s claims are ill-founded. Formal rule theories of propositional reasoning have 3 major weaknesses in comparison with the model theory: (a) They have no decision procedure; (b) they lack predictive power, providing no account of several robust phenomena (e.g., erroneous conclusions tend to be consistent with the premises); and (c) as a class of theories, they are difficult to refute experimentally.
There were three dominant reasoning research programs

Representational battle: Mental Logic  Mental Models  Heuristics and Biases
“The picture that emerges from this focus on mechanistic explanation is of the cognitive systems as an assortment of apparently arbitrary mechanisms, subject to equally capricious limitations, with no apparent rationale or purpose.”
-- Chater & Oaksford, 1999
Representational battle: Mental Logic Mental Models Heuristics and Biases

Probabilistic approaches
Representational battle:  Mental Logic  Mental Models  Heuristics and Biases

Pragmatists  Experimental Pragmatics  Probabilistic approaches
Representational battle: Mental Logic

Mental Models

Heuristics and Biases

Pragmatists

Experimental Pragmatics

Probabilistic approaches

Here we are today
One can superimpose Marr’s 3 levels on this too:

Computational
Algorithmic
Implementational
Representational battle: Mental Logic  

Mental Models  

Heuristics and Biases

Pragmatists  
Experimental Pragmatics (algorithmic & implementational)

Probabilistic approaches (computational)

Here we are today
See Love, 2015

“Focusing solely on the environment, numerous theoretical constraints are discarded, such as those provided by physiology, neuroimaging, reaction time, heuristics and biases, and much of cognitive development.”
Now, during that earlier epoch, the focus was on conditionals. When the computational approach came around, focus was on the fact that

a) The material conditional is not a good model,
b) modus ponens is distinguishable from modus tollens,
c) conditional inferences are defeasible....
d) conditionals lead to inferences differently as a function of negatives...

...etc. (see Oaksford & Chater).

Not much has been devoted to knowing more about conditional processing itself.

Compare this to scalars, which explored in detail the frontier between semantics and pragmatics of key terms such as some, or, might etc. (though primarily some)
When compared to scalars (a very small set of terms) – which has now been the main example of analyzing the semantics/pragmatics frontier -- this same frontier wrt conditionals has gotten only a cursory glance from experimentalists...In other words, there is actually a paucity of information about conditional processing per se.

There is still much more to know about conditionals before modeling them.

Before modeling conditionals, the semantics of conditionals would benefit from an experimental pragmatic treatment.
To summarize very briefly, the empirical literature has established that pragmatic enrichments, as exemplified by scalars (e.g. enriching an utterance with *Some* as *Some but not all*), are costly to generate based on the following sort of findings:

i) children enrich less often than adults

ii) scalars are *not* produced systematically among adults

iii) presence of scalar inference is linked with longer reading times (Bott & Noveck, 2004)

iv) sentence-processing and eye-tracking studies show that an enriched reading (of ‘Some’) does not immediately disambiguate between two options (Breheny et al., 2007; Huang & Snedeker, 2009)

v) Theory of mind abilities arguably figure into scalar inference making (Nieuwland et al., 2010)
From Pouscoulous, Noveck, Politzer & Bastide, (LA, 2007)
Percentage of children and adults who respond logically to 4 of the 10 critical statements presented in Experiment 1, including the one of central interest (in bold).

<table>
<thead>
<tr>
<th>Statements</th>
<th>Logical response</th>
<th>Children 9-10 y.o. (N=23)</th>
<th>Adults (N=19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All the turtles are in the boxes</td>
<td>True</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Some turtles are in the boxes</td>
<td>True</td>
<td>91%</td>
<td>53%</td>
</tr>
<tr>
<td>Some turtles are not in the boxes</td>
<td>False</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>All of the dolphins are in the boxes</td>
<td>False</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Giant COST A33 project initiated by Uli; Napoleon Katsos coordinated the investigation on scalars; Chris Cummins and I also played important roles...(just resubmitted)
What these examples have in common is a gain in informativity, where the pragmatic inference reduces the number of possible true cases.
Adult processing tells a similar story
Following up on a developmental study showing children’s reticence to enrich scalar utterances (Noveck, 2001), consider a categorization task where the quantifier is weaker than it ought to be:

For example,
Some cows are mammals (Certaines vaches sont des mammifères).
Some trout are fish (Certaines truites sont des poissons).
Some parakeets are birds (Certains perroquets sont des oiseaux).

Of course...
All cows are mammals.
All trout are fish.
All parakeets are birds.

Thus, an interlocutor has a reason to be tempted to say that "Some cows are mammals" is false because it implicates that Not All cows are mammals.
Lewis Bott and I (Bott & Noveck, 2004, JML) ran 4 experiments based on the following paradigm:

54 items randomly presented by computer.

6 Categories -- mammals, fish, reptiles, shellfish, birds, fruit

1) Some cows are mammals. * (True logically/False with enrichment)
2) Some mammals are cows. (True)
3) Some cows are insects. (False)
4) All cows are mammals. (True)
5) All mammals are cows. (False)
6) All cows are insects. (False)
Experiment 3: Whole sentences, no specific instruction

*e.g. Some cows are mammals (Certaines vaches sont des mammifères)*

Presented 54 items (6 categories; 9 per condition) and asked 32 participants to respond "True" or "False".

41% True and 59% False (in line with other data; Noveck, 2001)
T1) Some cows are mammals.  (Logic - true / Pragmatic - false)
T2) Some mammals are cows.  (True)
T3) Some cows are insects.  (False)
T4) All cows are mammals.  (True)
T5) All mammals are cows.  (False)
T6) All cows are insects.  (False)
Experiment 4: Controlling the amount of available effort.

The same task as before but with the following features

- One word at a time (to control uptake)

- Two lag times:  
  A) Short amount of time (900 msecs)  
  B) Long amount (3 seconds).
Percentage saying "true" to each of the Sentence types in Experiment 4. N= 45.

<table>
<thead>
<tr>
<th>Utter.</th>
<th>Example</th>
<th>Short Lag (900 msecs)</th>
<th>Long lag (3000 msecs)</th>
<th>Logical response diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>Some robins are birds</td>
<td>.72</td>
<td>.56</td>
<td>-.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(p &lt; .001)</td>
</tr>
<tr>
<td>T2</td>
<td>Some birds are robins (T)</td>
<td>.79</td>
<td>.79</td>
<td>.00</td>
</tr>
<tr>
<td>T3</td>
<td>Some robins are fish    (F)</td>
<td>.12</td>
<td>.09</td>
<td>+.03</td>
</tr>
<tr>
<td>T4</td>
<td>All robins are birds    (T)</td>
<td>.75</td>
<td>.82</td>
<td>+.07</td>
</tr>
<tr>
<td>T5</td>
<td>All birds are robins    (F)</td>
<td>.25</td>
<td>.16</td>
<td>+.09</td>
</tr>
<tr>
<td>T6</td>
<td>All robins are fish     (F)</td>
<td>.19</td>
<td>.12</td>
<td>+.07</td>
</tr>
</tbody>
</table>
Bimodal distributions are generally the rule.

• Strategies remain largely consistent throughout a session.

• Situational factors such as quickened response times, or secondary tasks, can prevent deeper processing.

• Individual differences figure into prompting different interpretations.
These were all findings before computational accounts came into the picture.
1) One cannot generalize from theoretical approaches to scalars...

On the one hand, enrichments of weak utterances appear to restrict meanings...

Conditionals on the other hand, do not....
2) The conditional is...

- Well known for its pragmatic potential...

Geis & Zwicky (1971) in their seminal squib:

[This] regular association [linking *if* and the situations in which they arise]....asserts a connection between linguistic form and a tendency of the human mind ‘to perfect conditionals to biconditionals’ in words suggested to us by Lauri Kartunnen. This tendency is manifested in two classical logical fallacies, Affirming the Consequent (concluding X from $X \supset Y$ and $Y$) and Denying the Antecedent (concluding $\sim Y$ from $X \supset Y$ and $\sim X$).
There are many inference forms associated with conditionals

**Two pragmatically justified ones**

**Affirmation of the Consequent**

If Jean goes to the cinema, then he travels by bicycle  
Jean travels by bicycle.  
Therefore, he goes to the cinema

**Denial of the Antecedent**

If Jean goes to the cinema, then he travels by bicycle  
Jean does not go to the cinema  
Therefore, he does not travel by bicycle
3) Invited inferences vary quite a bit (like with scalars)

Here’s an example of a low rate of AC endorsements

![Bar chart showing endorsement rates for Modus Ponens and Affirmation of Consequent.]

e.g. Marcus & Rips, 1975
Highest rate of endorsement – 79%

(e.g. Barrouillet et al. 2000)
4) An fMRI study (Noveck, Goel, & Smith, 2004) that focused mostly on the valid inference forms, Modus Ponens & Modus Tollens.
Modus Ponens - Baseline

Modus Tollens - Baseline

N=16
A) We captured little or no activity for these two forms in the conclusion.

B) Rather high rates of correct responses (saying Inconclusive).

-- 69% said inconclusive to AC
-- 60% said inconclusive to DA
5) Existing accounts:

Break them down into two groups:

I. Predominant group: Those that encourage AC interpretations

1. Horn: If is strengthened directly into iff: likens it to cases such as “drink” for alcoholic beverage or how “I don’t believe that p” becomes “I believe that not-p”

2. Braine and Mental Logic: If $p$ then $q$ comes with an invited inference “If $q$ then $p$."

...
II. Those that discourage an AC interpretation

1. Markovits. If P then Q (at least for causals) prompts If A then Q as in
   *If the electricity goes out then school is cancelled* prompts other causes, such as snow.

   \[ \text{If } A \text{ then } Q \]
   
   If P then Q
   Q

2. Von Fintel:

   \[ Q \text{ (no matter what)} \]
   
   If P then Q

   If that is the case, then
   The minor premise in

   \[ \text{If } P \text{ then } Q \]
   Q
   Not Q

   prompts a contradiction
1) We used an Experimental Pragmatic strategy of isolating pragmatic effects in order to better understand the lexical contribution made by conditionals.

2) We dealt only with Modus Ponens and Affirmation of the Consequent; we avoid negations.

3) We use temporal measures (reading time measures and EEG).
With an eye toward an ERP experiment...

Modus Ponens

If John watches television then he eats cookies.
He watches television.
+ He eats cookies.

MP’: + He eats fries.

Affirmation of the consequent

If John takes a metro then he reads a newspaper.
He reads a newspaper.
+ He takes a metro.

AC’: + He takes a bicycle.
Si Jean regarde la télé alors il mange des gâteaux.
Il regarde la télé.

MP  Il mange des gâteaux.

MP'  Il mange des frites.

Si Jean regarde la télé alors il mange des gâteaux.
Il mange des gâteaux.

AC  Il regarde la télé.

*AC'  Il regarde la lune.
Experiment with the MP/MP’/AC/AC’ paradigm:

Behavioral study

% correct responses

Slowing down is linked to an increased likelihood to respond “normatively” to AC syllogisms....

Van der Henst, Ciceron, Bujakowska & Noveck, 2006
Modus Ponens vs Affirmation of the Consequent: First word in the minor premise indicating whether MP or AC is called for (Word 12)

With respect to the Minor Premise of Modus Ponens, The AC Minor premise is linked to an N400, an indication of inconsistency.

Difference between MP & AC: 200-300 ms
If P then Q
P

Versus

P
If P then Q

Fig. 1. Experimental design. Top: minor premise presented after the conditional statement; bottom: minor premise presented before the conditional statement. The numbers correspond to the number of trials in each condition.
The minor premise of a Modus Ponens conditional elicits a P3b, an indication that an expectation was satisfied.
Our new, sparer paradigm...

Bonnefond et al. (JML, 2012)
Main interest: How does the second premise react, even as we expect two sorts of responses...

That is, we expect two groups in response to AC: Endorsers and Rejectors

**Affirmation of the Consequent**

If P then Q

\[
\begin{array}{c}
Q \\
P
\end{array}
\]

Accept as “logical”

**Modus Ponens**

If P then Q

\[
\begin{array}{c}
P \\
Q
\end{array}
\]

Accept as “not-logical”

If the conditional comes with an enrichment that *facilitates* an AC inference, then the reaction to the Minor premise should be like that of Modus Ponens...

If the conditional comes with an enrichment that *discourages* an AC inference, then the reaction to its Minor premise should be slower than the one in Modus Ponens...
Two studies: One behavioral and one EEG
If there is a J then there is a N.
There is a J.
There is a N.
Premise 1

<table>
<thead>
<tr>
<th></th>
<th>If $P$ then $Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Endorsers 1805</td>
</tr>
<tr>
<td></td>
<td>Mixed 2440</td>
</tr>
<tr>
<td></td>
<td>Rejectors 2552</td>
</tr>
</tbody>
</table>

Premise 2

<table>
<thead>
<tr>
<th></th>
<th>$P$ (MP)</th>
<th>$Q$ (AC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Endorsers 812</td>
<td>Endorsers 854</td>
</tr>
<tr>
<td></td>
<td>Mixed 875</td>
<td>Mixed 1075</td>
</tr>
<tr>
<td></td>
<td>Rejectors 987</td>
<td>Rejectors 1277</td>
</tr>
</tbody>
</table>

Conclus’n

<table>
<thead>
<tr>
<th></th>
<th>$Q$ (MP)</th>
<th>$R$ (MP)</th>
<th>$P$ (AC)</th>
<th>$S$ (AC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Compatible</td>
<td>Incompatible</td>
<td>Compatible</td>
<td>Incompatible</td>
</tr>
<tr>
<td></td>
<td>Endorsers 847</td>
<td>965</td>
<td>$835^e$</td>
<td>892</td>
</tr>
<tr>
<td></td>
<td>Mixed 843</td>
<td>925</td>
<td>$1000^e/888^r$</td>
<td>921</td>
</tr>
<tr>
<td></td>
<td>Rejectors 966</td>
<td>1131</td>
<td>$1230^r$</td>
<td>1055</td>
</tr>
</tbody>
</table>

Styles (n) = Endorsers (15), Rejectors (13) & Mixed (9)
EEG study -- What to look for:

• Second Premise

• N2 indicates a violation
• P3b indicates an expectation that was met

• Focus on Endorsers and Rejectors
Second Premise **Rejectors**

![Graphs showing N2 and P3b waveforms across different brain regions (Frontal, Central, Centroparietal, Parietal). The graphs display waveforms for two conditions: MP and AC.](image)
Everyone rejects AC’s second premise; everyone arguably expects MP’s minor premise

-- what an individual will do in light of that will vary
-- the distinction between MP & AC is more pronounced among the Rejectors
### Varieties of narrowing

<table>
<thead>
<tr>
<th>Voluntary</th>
<th>Coerced</th>
</tr>
</thead>
<tbody>
<tr>
<td>The drawing out of a pragmatic inference is optional</td>
<td>The drawing out of a pragmatic inference is obligatory in order to make sense of the utterance</td>
</tr>
<tr>
<td>The (hypothesized) intended reading relies on a narrower version of the linguistically encoded one</td>
<td>There is no obvious relationship between the linguistically encoded reading and the (hypothesized) intended one</td>
</tr>
<tr>
<td>The linguistically encoded reading is often good enough for discerning a speaker's informative intention</td>
<td>The linguistically encoded meaning is implausible</td>
</tr>
<tr>
<td>When the linguistically encoded reading is narrowed, extra effort is generally detectable</td>
<td>Extra effort compared to what?</td>
</tr>
</tbody>
</table>

*“invited inferences” go here*
Does this distinction allow us to cash anything else out?

Yes

A lot of work shows that cases – that are arguably the voluntary kind – are not problematic for ASD folks (e.g. Chevallier, Happé, Wilson & Noveck, 2010), but there are differences of the coerced kind.
Scalar Inferences in Autism Spectrum Disorders

Caroline Chevallier - Nicole Wilson - Francesco Happé - Ted Novick

Table 3: Percentage of correct answers as a function of connective type (And, Or), condition (FF, FT/TF, TT) and group (TD, ASD); results of Mann-Whitney U tests comparing the groups TD (N = 22) and ASD (N = 22).

<table>
<thead>
<tr>
<th>Connective</th>
<th>Truth condition</th>
<th>TD (N = 22) Mean (SD)</th>
<th>ASD (N = 22) Mean (SD)</th>
<th>Mann-Whitney U test</th>
<th>Difference in mean ranks (Effect size index)</th>
</tr>
</thead>
<tbody>
<tr>
<td>And</td>
<td>FF</td>
<td>100 (0)</td>
<td>99 (4)</td>
<td>z = 0.00; p = 1.00</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>TF-FT</td>
<td>99 (4)</td>
<td>99 (4)</td>
<td>z = -0.77; p = 0.44</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>TT</td>
<td>100 (0)</td>
<td>97 (7)</td>
<td>z = -0.25; p = 0.80</td>
<td>-0.02</td>
</tr>
<tr>
<td>Or</td>
<td>FF</td>
<td>42 (45)</td>
<td>48 (48)</td>
<td>z = 0.83; p = 0.40</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>TF-FT</td>
<td>43 (44)</td>
<td>57 (47)</td>
<td>z = 0.35; p = 0.72</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>TT</td>
<td></td>
<td></td>
<td>z = 0.01; p = 0.99</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Arguably, ASD deficits are linked to the coercion label.
Summary:

Using scalars as an experimental model for analyzing the semantic/pragmatic frontier, we investigated conditionals (at that same frontier).

Conditionals (at least *out of the blue* conditionals in our restrained environment), show that listeners anticipate MP inferences over AC inferences.

Conditionals do not fit into scalar (and referential) framework. There is a distinction to be made (between what I call *voluntary* and *coerced* cases). This distinction allows one to cash out empirical findings.
Danke