Nonmonotonic logic as a cognitive tool

Cognitive Science Center Amsterdam

Michiel van Lambalgen

Psychology and the tyranny of the paradigm

- psychology is dominated by experimental paradigms or tasks
- paradigms take a great deal of investment to establish
- and so it is hoped the data they furnish will be criterial for some theoretical concept
- not surprisingly, matters usually turn out somewhat more complicated
- our proposal is that logic, sensitively applied, can help to relate paradigms/tasks in a more explicit way
- examples: the suppression task, the false belief task, the box task, the tubes task, the counterfactual reasoning task, . . .

Nonmonotonic logic

- logic programing and closed world reasoning
- why this logic?
- biographically, because we both come from discourse reasoning/interpretation/semantics
- biologically, because we believe it is appropriate for capturing a cognitively central capacity
- updating a model of the situation as a basis for action, in the light of incoming information, using general knowledge of regularities, and particular context, nonlinguistic or linguistic
- it is very nearly the opposite of classical logic
- there can be no doubt that many logics are needed for modelling cognition

The Suppression Task: defeasible discourse interpretation

The discourse starts, e.g.: "If she has an essay to write she will study late in the library"
and then continues with a categorial premiss:
e.g., "She has an essay to write"
What follows?
Or there may be extra conditional premisses in between the first, and the categorial premiss:
'Additional' Premises (abnormalities): "If the library is open, she'll be in the library"
'Alternative' Premises: (other motives): "If she has a textbook to read, then she'll be in the library"

Nonmonotonic analysis

- $P \wedge \neg ab \rightarrow Q$ (e.g. If she has an essay and nothing is abnormal, then she'll be in the library)
- the 'closed-world-assumption' : if we can't prove a proposition from what is in the discourse up to here, then its negation is true e.g. $\neg ab$, at this point
- "The library is closed" is an abnormality relative to this conditional
- NB in the suppression task, an indirect conditional 'hint' puts the abnormality into play: *'If the library is open, she's in the library'*

The Suppression Task: formal structures

Four possible conditional inference patterns:

$p \to q, p, \vdash q$: *modus ponens* (MP)

$p \to q, \neg p \vdash \neg q$: *denial of the antecedent* (DA)

$p \to q, \neg q \vdash \neg p$: *modus tollens* (MT)

$p \to q, q \vdash p$: *affirmation of the consequent* (AC)

MP and MT are 'classically valid'; DA and AC are not.

DA and AC may be nonmonotonically valid, and MP and MT nonmonotonically invalid, depending on what else is in the database/discourse as so far developed.

- e.g. If she has an essay, then she's in the library. The library is closed. (MP doesn't apply)
- e.g. If she has an essay, she's in the library. She's in the library. (AC does apply, until . . .) If she has a textbook to read, she's in the library.

The Suppression Task: materials

- If she has an essay to write she will study late in the library.
  She has an essay to write
  So . . . ?
- If she has an essay to write she will study late in the library.
  She will study late in the library.
- If she has an essay to write she will study late in the library.
  She doesn't have an essay to write
- If she has an essay to write she will study late in the library.
  If the library stays open she will study late in the library.
  She has an essay to write
- If she has an essay to write she will study late in the library.
  If the library stays open she will study late in the library.
  She will study late in the library

The Suppression Task — materials (cont.)

- If she has an essay to write she will study late in the library.
  If the library stays open she will study late in the library.
  She doesn't have an essay to write

- If she has an essay to write she will study late in the library.
  If she has some textbooks to read, she will study late in the library.
  She has an essay to write.

- If she has an essay to write she will study late in the library.
  If she has some textbooks to read, she will study late in the library.
  She does not have an essay to write.

- If she has an essay to write she will study late in the library.
  If she has some textbooks to read, she will study late in the library.
  She will study late in the library.

The suppression task in 'autists' as compared to 'normals'

	ASD			Control		
% responses	**yes**	**no**	**maybe**	**yes**	**no**	**maybe**
MP	89.6	0.0	10.4	96.1	2.5	1.4
MP add	71.0	1.1	28.0	51.1	0.7	48.2
MP alt	92.9	0.4	6.8	97.5	0.7	1.8
MT	1.4	79.6	19.0	2.5	92.8	4.7
MT add	0.7	62.1	37.1	0.7	45.0	54.3
MT alt	0.4	90.3	9.3	1.1	95.0	3.9
AC	45.0	1.1	53.9	67.1	2.1	30.7
AC add	28.1	1.1	70.9	35.7	0.0	64.3
AC alt	12.2	2.2	85.7	9.6	0.0	90.4
DA	1.1	48.0	50.9	0.4	69.1	30.6
DA add	2.9	28.9	68.2	2.5	33.6	63.9
DA alt	3.2	15.7	81.1	1.1	10.4	88.5

Table 5. Proportion of responses for the simple task and the suppression task.
add = additional and alt=alternative premise.

The false belief task (Wimmer & Perner)

- doll (Maxi) and child witness chocolate placed in a box
- exit Maxi
- the child sees the experimenter move chocolate from box to drawer
- re-enter Maxi
- child is asked 'where will Maxi look for the chocolate?'
- before about 4 yrs, child responds where child knows chocolate to be
- after 4 yrs, child responds where Maxi falsely believes chocolate to be
- autists go on failing this task for a long time

Sketch of a logical analysis of the false belief task

- the 'database' contains response rules of the form

$$B(\varphi) \wedge \neg ab_{\varphi} \rightarrow R(\varphi) \tag{1}$$

- 'if an agent *B*elieves φ and nothing *ab*normal is the case, then the agent *R*eports φ'
- two possibly competing instances of 1
  - let φ represent the actual location of the chocolate
  - let ψ represent Maxi's belief about the location of the chocolate, $\psi = B_m(\theta)$
  - here $\varphi \rightarrow \neg\theta$
- the two response rules

$$B(\varphi) \wedge \neg ab_{\varphi} \rightarrow R(\varphi); B(\psi) \wedge \neg ab_{\psi} \rightarrow R(\psi) \tag{2}$$

must inhibit each other, since only one answer can be given

Sketch of a logical analysis of the false belief task

- inhibition proceeds via the abnormalities, e.g. with the same notation as above

$$B(B_m(\theta)) \rightarrow ab_\varphi \qquad (3)$$

- this clause reflects task understanding: 'if Maxi believes something that is false, the agent should suspend his own response based on true belief'
- task understanding must thus inhibit the response – if inhibition is compromised/immature, or if the executive rule 1 has no slot for an abnormality, the child may give its own response, φ
- which in turn inhibits the ψ-response via $R(\varphi) \rightarrow ab_\psi$
- for the computation that leads to $B(B_m(\theta))$ see next two slides
- the formal structure is analogous to the suppression task: we predicted that autistic subjects would have trouble with abnormalities in the suppression task

Inertia and closed world reasoning, and the FBT

- pragmatically, the formulation of the 'false belief' question suggests it must have an answer
- classical logic compels one to ask 'what else could be the case?', reflecting the obligation to consider *all* models of the data,
- nothing of the sort happens in actual causal reasoning; instead
  1. one assumes that only those events (affecting the entity of interest) occur which are forced to occur by the data – here the only such event is the chocolate's change of location from box to drawer
  2. one also assumes that events only have those causal effects which are described by one's background theory
  3. no spontaneous changes occur, that is, every change of state or property can be attributed to the occurrence of an event with specified causal influence

Attribution of beliefs and closed world reasoning

- awareness of the causal relation between perception and belief, which can be stated in the form: 'if φ is true in scene S, and agent a sees S, then a comes to believe φ'
- thus Maxi comes to believe that the chocolate is in the box
- the principle of inertia (cf. 3 above) yields that Maxi's belief concerning the location of the chocolate persists unless an event occurs which causes him to have a new belief, incompatible with the former
- the story does not mention such an event, whence it is reasonable to assume – using 1 and 2 – that Maxi still believes that the chocolate is in the box when he returns to the experimenter's room
- attribution of belief is thus a special case of causal reasoning, but the causal relation between perception and belief is an essential ingredient in the false belief task, absent in the counterfactual task

The lesson of nonverbal false belief tasks

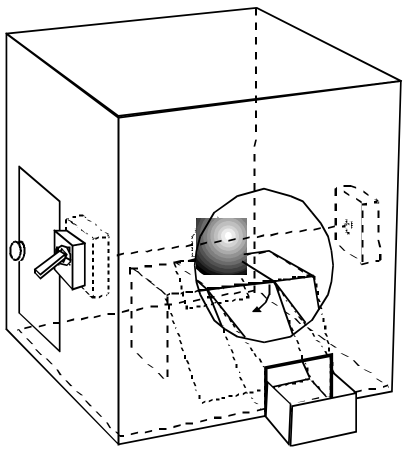
- Clements and Perner 1994
  - eye gaze response is measured after experimenter's *prompt*: 'I wonder where Sam mouse will look for his cheese'
  - verbal/gestural response measured after experimenter's *question*: 'Where will Sam mouse look for his cheese?'
  - from age 2;11 till age 4;6 the eye gaze response is more accurate than the verbal/gestural response
  - from 3;8 till 4;6 eye gaze is 100% accurate, verbal/gestural response only in 63%
- authors' explanation in terms of logical notions

'[There is] a distinction between representing a fact and making a judgement about that fact. Answering a question, whether verbally or by pointing, requires a judgement. Looking in anticipation does not. [...] A single mental model is sufficient to represent the current state of the world and to act on it. That is, pure action (i.e., looking in anticipation) is done only on the basis of a representation of reality; that is, one model. But to make a judgement (verbally or gesturally) at least two models are required: One to represent the proposition to be judged (information expressed), and the other to represent the state of the world by which this proposition is to be judged.

The meaning of nonverbal false belief tasks?

- ‘In other words, to make a judgement is to convey that the verbally or otherwise expressed information (the model of whatever is being proposed) conforms with reality (the other model).’
- but taken too literally, this could lead to regress
- it points to the need to elaborate executive functions—the development of capacities to monitor, to suppress alternatives, working memory (for end-result vs. process), . . .
- and the ‘two models’ vs. ‘one model with ‘post-its’ ’ question is recurrent: reminiscent of adult deductive reasoning issues

The box task [1] (Russell)



- Russell's 'box task': to get the marble you have to flip the switch first, or the marble drops out of reach when the light-beam is broken
- instructions to subject involve pointing this out *explicitly*

Russell's comments

"[T]aking what one might call a 'defeasibility stance' towards rules is an innate human endowment – and thus one that might be innately lacking . . . [H]umans appear to possess a capacity – whatever that is – for abandoning one relatively entrenched rule for some novel ad hoc procedure. The claim can be made, therefore, that this capacity is lacking in autism, and it is this that gives rise to failures on 'frontal' tasks – not to mention the behavioral rigidity that individuals with the disorder show outside the laboratory. . . "

Russell goes on to say that one way this theory might be tested is through the implication that "children with autism will fail to perform on tasks which require an appreciation of the defeasibility of rules such as 'sparrows can fly'."

Performance and closed world reasoning in the box task

- prepotent response is to reach for the marble, which must be inhibited to achieve correct response
- (normal) children older than 4yrs manage this, younger children don't
- hypothesis: in box task, the 'qualification problem' (subsidiary of the 'frame problem') is at work – in general one cannot list beforehand all preconditions of an action (hence one has to be prepared to adapt to circumstances)
- normally developing children learn to interpret conditional as allowing for exceptions: the general form is

$$p \wedge \neg ab \rightarrow q,$$

  where conditions $\neg r \rightarrow ab$ may be added; then apply closed world reasoning to ab

Formal model of reasoning in the box task: language

For the formalization we borrow some self-explanatory notation from the situation calculus. Let c be a variable over contexts, then the primitives are

– the predicate $do(a, c)$, meaning 'perform action a in context c'

– the function $result(a, c)$, which gives the new context after a has been performed in c.

The actions we need are g ('grab'), u ('switch up'), d ('switch down'). We furthermore need the following context-dependent properties:

– $possess(c)$: the child possesses the marble in c

– $up(c)$: the switch is up in c ($=$ correct position)

– $down(c)$: the switch is down in c ($=$ wrong position).

Formal model of reasoning in the box task: principles

The following equations give the rules appropriate for the box task

$$down(c) \wedge do(u,c) \wedge \neg ab'(c) \rightarrow up(result(u,c)) \tag{4}$$

$$do(g,c) \wedge \neg ab(c) \rightarrow possess(result(g,c)) \tag{5}$$

We first model the reasoning of the normal child $>$ 4 yrs. Initially, closed world reasoning for $ab(c)$ gives $\neg ab(c)$, reducing the rule 5 to

$$do(g,c) \rightarrow possess(result(g,c)) \tag{6}$$

which prompts the child to reach for the marble without further ado. After repeated failure, she reverts to the initial rule 5, and concludes that after all $ab(c)$. After the demonstration of the role of the switch, she forms the condition

$$down(c) \rightarrow ab(c) \tag{7}$$

Formal model of reasoning in the box task: normal reasoning

She then applies closed world reasoning for ab to 7, to get

$$down(c) \leftrightarrow ab(c) \tag{8}$$

which transforms rule 5 to

$$do(g, c) \wedge up(c) \rightarrow possess(result(g, c)) \tag{9}$$

Define context c_0 by putting $c = result(u, c_0)$ and apply closed world reasoning to rule 4, in the sense that $ab'(c)$ is set to $\bot$ due to lack of further information, and $\rightarrow$ is replaced by $\leftrightarrow$. Finally, we obtain the updated rule, which constitutes a new plan for action

$$down(c_0) \wedge do(u, c_0) \wedge c = result(u, c_0) \wedge do(g, c) \rightarrow possess(result(g, c)) \tag{10}$$

Formal model of reasoning in the box task: autistic reasoning

The normal child < 4 yrs and the autistic child are assumed to operate effectively with a rule of the form

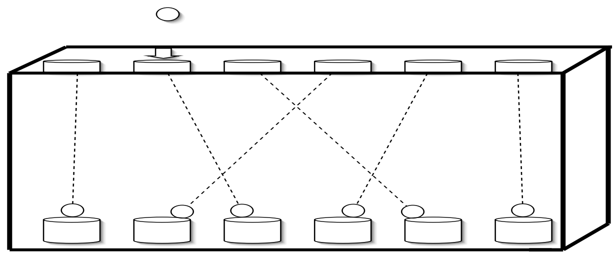
$$do(g, c) \rightarrow possess(result(g, c)) \tag{11}$$

which cannot be updated, only replaced *in toto* by a new rule such as 10.

Jim Russell wrote that ‘humans appear to possess a capacity – whatever that is – for abandoning one relatively entrenched rule for some novel ad hoc procedure’. The preceding considerations suggest that ‘abandoning one relatively entrenched rule’ may indeed be costly, but that normal humans get around this by representing the rule in such a way that it can be easily updated.

In normals, the only costly step appears to be the synthesis of the rule 7; recall the suppression task!

Russell's [3]'tubes task': a counterexample?



- child has to retrieve the ball in the appropriate tray below
- neither normal nor autistic child has difficulty doing this, i.e. in switching between rules
- but isn't this also a case of having to inhibit a prepotent response?
- Russell's first suggestion: there is a difference between natural ('tubes') and artificial ('box') default rules
- we will try to find a difference in *logical structure* of the tasks

Why do autists perform normally in the 'tubes task'?

- the rules involved in the two tasks (tubes, box) have different *logical forms*
- in the box task, correct performance hinges on the ability to amend the *antecedent* of the rule, whereas in the case of the tubes task it is the consequent (i.e. the catch-tray) that has to be changed
- in the box task, the original plan has to be changed by incorporating another action conjunctively, whereas in the tubes task the change is 'additive'
- the single action 'look in tray directly beneath hole' becomes
- a single IF-THEN-ELSE rule, where the action to be taken depends on the satisfaction or non-satisfaction of an explicit precondition:

  IF unimpeded fall of the ball THEN look in tray directly beneath hole ELSE follow tube

On the comparative ease of IF-THEN-ELSE processing

- Autists are good at IF-THEN-ELSE, as in the following typical experiment:
  - subjects were shown different letters of the alphabet that flashed one at a time on a computer screen, and were asked to respond by pressing a key in every case except when they saw the letter X
  - the first task was a Go task, in which the letter X never appeared and in this way subjects were allowed to build up a tendency to respond
  - immediately afterward, subjects performed a Go/NoGo task in which the letter X did appear in the lineup, at which point the subject had to control the previously built-up impulse to respond
- shows that autists first learn the IF-THEN-ELSE rule, then apply it
- the rule in the box task can neither be synthesized nor be learned

Summing up

- all these tasks have been analysed using 'logic programming with negation as failure', a particular nonmonotonic logic (also known as 'planning logic' and used in robotics)
- Generalities:
  - this logic is a model of fast, automatic, knowledge-rich reasoning *to* an interpretation of task materials
  - it implies some management' by 'executive functions' (inhibition of prepotent responses, handling model relations, ...
  - it is quite unlike classical logic (nonmonotonic, trivalent, knowledge-rich, produces single minimal models, ...) and is not particularly linguistic, even though it is mostly modelling discourses here

Summing up: Specifics

- logical analysis finds analogies between linguistic and nonlinguistic tasks (e.g. box/suppression tasks) and disanalogies between non-linguistic tasks (e.g. box/tubes tasks)
- can also find analogies and disanalogies between the same pair of tasks—as we shall see tomorrow
- and has been used to make predictions of behaviour (e.g. autistic children) from their behaviour on one task to their behaviour on another superficially highly dissimilar task (e.g. box/supression tasks) [2]

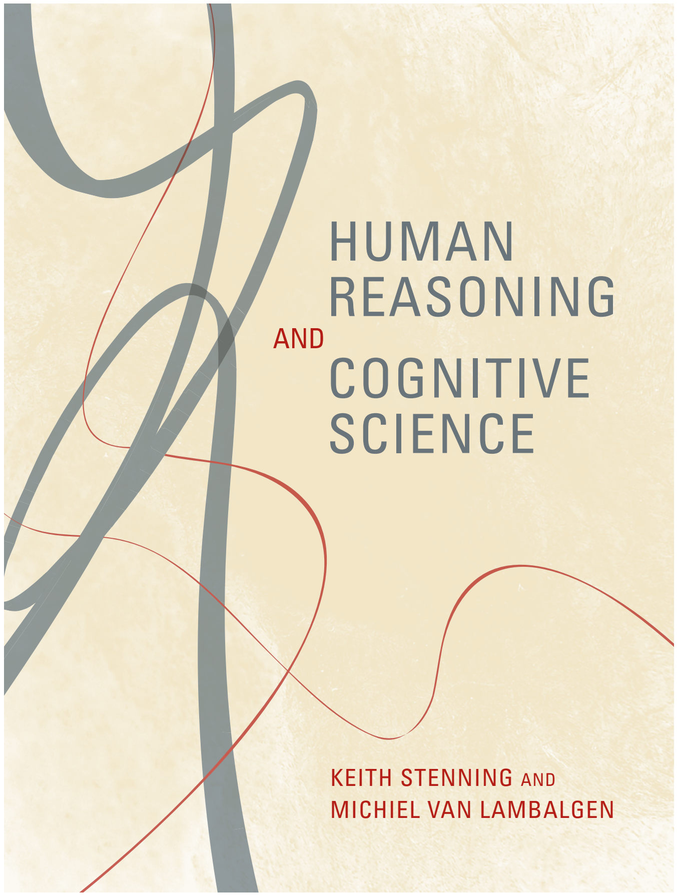

HUMAN
REASONING
AND
COGNITIVE
SCIENCE
KEITH STENNING AND
MICHIEL VAN LAMBALGEN

Figure 1: [4]

References

[1] C. Hughes and J. Russell. Autistic children's difficulty with disengagement from an object: its implications for theories of autism. *Developmental Psychology*, 29:498–510, 1993.

[2] Judith Pijnacker, Bart Geurts, Michiel van Lambalgen, Jan Buitelaar, and Peter Hagoort. Exceptions and anomalies: An ERP study on context sensitivity in autism. *Neuropsychologia*, 48(10):2940–2951, 2010.

[3] James Russell, Christopher Jarrold, and Bruce Hood. Two intact executive capacities in children with autism: Implications for the core executive dysfunctions in the disorder. *JOURNAL OF AUTISM AND DEVELOPMENTAL DISORDERS*, 29(2):103–112, 1999.

[4] K. Stenning and M. van Lambalgen. *Human reasoning and cognitive science*. MIT University Press, Cambridge, MA., 2008.