Psychological Research (2003) 67: 219–231
DOI 10.1007/s00426-002-0092-7

ORIGINAL ARTICLE

Frank Rijmen · Paul De Boeck

A latent class model for individual differences in the interpretation of conditionals

Received: 22 August 2000 / Accepted: 20 March 2002 / Published online: 8 March 2003
© Springer-Verlag 2003

Abstract We investigated the hypothesis that there are three levels of performance associated with conditional reasoning: (1) Unsophisticated reasoners solve a modus tollens by accepting the invited inferences, treating the conditional as if it were a biconditional. (2) Reasoners of an intermediate level can resist the invited inferences, but cannot find the line of reasoning needed to endorse modus tollens. (3) Sophisticated reasoners do not draw the invited inferences either, but they do master the strategy to solve a modus tollens.

On a first set of six problems, solved by 214 adolescents, an unrestricted latent class analysis revealed the existence of a large subgroup of reasoners with a biconditional interpretation of the conditional, and a smaller subgroup with a conditional interpretation.

On a second set of 24 problems, solved by the same participants, a restricted latent class model corroborated the existence of a large subgroup of unsophisticated reasoners and a smaller subgroup of reasoners of an intermediate level. No evidence was found for the existence of a subgroup of sophisticated reasoners.

As expected, the class of biconditional reasoners was associated with the class of unsophisticated reasoners, and the class of conditional reasoners was associated with the class of reasoners of an intermediate level. Furthermore, the former showed a biconditonal response pattern on truth table tasks, whereas the latter showed a conditional response pattern.

Reasoning with conditionals (if/then propositions) is a central topic in the study of deductive reasoning. Four classical conditional reasoning problems are: modus ponens (MP; "if p, then q" and "p", hence "q"), modus tollens (MT; "if p, then q" and "not q", hence "not p"), denial of the antecedent (DA; "if p, then q" and "not p", hence "not q") and affirmation of the consequent (AC; "if p, then q" and "q", hence "p"). The latter two are fallacies; they are not valid arguments in classical logic.

The empirical studies of reasoning with conditionals indicate that MP is endorsed by almost everyone (except very young children), in contrast with MT (for a review, see Evans, Newstead, & Byrne, 1993). *Rule theories* of propositional reasoning (e.g. Braine, 1978; Braine, Reiser & Rumain, 1984; Braine & O'Brien, 1991; Osherson, 1975; Rips, 1984, 1994) account for this empirical finding by assuming that a rule for MP is part of the repertory, but a rule for MT is not. Hence, solving a MT problem should be more difficult because it requires the application of several basic rules:

- p (supposition)
- q (modus ponens)
- incompatible (contradiction)
- not p (reductio ad absurdum)

Also the *mental model theory* of propositional reasoning (Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne & Schaeken, 1992) can account for the difference between MP and MT. MP can be solved using the initial representation of a conditional:

- [p] q

so that MP requires only one explicit model, whereas MT requires that the initial representation is fleshed out to represent all three of the situations that are consistent with the conditional:

- p q
- not p q
- not p not q

Another empirical observation is that participants often do endorse the DA and AC inferences, although they are not valid in classical logic. Avoiding the

Frank Rijmen was supported by the Fund for Scientific Research Flanders (FWO).

F. Rijmen (✉) · P. De Boeck
University of Leuven, Tiensestraat, 102, 3000 Leuven, Belgium
E-mail: frank.rijmen@psy.kuleuven.ac.be
Tel.: +32-16-326123
Fax: +32-16-325988

incorporation of rules for these invalid inferences in the repertory, rule theories typically ascribe these inferences to the phenomenon of invited inferences (Geis & Zwicky, 1971). For example, a conditional "if p, then q" often invites the inferences "if not p, then not q", and "if q, then p" (treating the conditional as if it were a biconditional). Though these invited inferences do not follow logically from the conditional, they are often implied in real-world conversations. For example, the statement "If you mow the lawn, I'll give you five dollars" probably means that I will not give you five dollars if you don't mow the lawn. Following the "logic of conversation", many participants accept these invited inferences unless they have reason to believe them to be inappropriate. One way to countermand the invited inferences is by expanding the major premise (Rumain, Connell & Braine, 1983; Experiment l):

- If the bottom has *N*, then the top has *X*

as follows:

- But if the bottom doesn' t have *N*, then the top may have *X* or it may have some other letter. And, if the top has *X*, then the bottom may have *N* or it may have some other number.

Adding the additional premises emphasizes the possibility of alternative conditions ("some other number") for the consequent, and, by consequence, blocks the invited inferences.

A remarkable effect of expanding the major premise was that, apart from the expected decrease in the endorsement of the fallacies DA and AC, the endorsement of the logical valid MT inference also decreased. The authors explain this by assuming that there are two ways to solve a MT: a primitive one, which consists in accepting the invited inference, treating the conditional as a biconditional, and a more sophisticated one, which is the reductio ad absurdum (see example above). The decrease in performance on MT problems is due to the blocking of the invited inferences and the corresponding primitive routine. In line with this proposed explanation, Johnson-Laird et al. (1992) found that MT is indeed less endorsed with conditionals than with biconditionals. Hence, MT is more difficult under a pure conditional interpretation.

In addition to the hindering effect of a conditional interpretation on the endorsement of MT, Rumain et al. (1983) go one step further in their explanation, asserting that "the decrease in 'no' responding [the decrease in performance on MT problems] when an expanded premise is used marks an *upward shift*, we believe, in the *sophistication of reasoning*" (p.475, italics and comment added). Hence, Rumain et al. (1983) suggest that the way a conditional is interpreted is related with general (propositional) reasoning ability. They propose *three levels of performance*: (1) Unsophisticated reasoners (e.g. children) will solve a MT by accepting the invited inferences. (2) Reasoners of an intermediate level can resist the invited inferences, but cannot find the reductio ad absurdum line of reasoning needed to endorse MT. (3) Sophisticated reasoners do not draw the invited inferences either, but they do master the reductio-strategy and hence can perform a MT inference (see also Braine et al., 1984; Braine & O'Brien, 1991). Expanding the major premise brings a reasoner from the first level to the second one for the problems that are presented.

It should be noted that this level idea, where interpretation of the conditional and propositional reasoning ability are related, also fits within the framework of the mental model theory: Unsophisticated reasoners, in fleshing out the initial representation, consider only those models where the antecedent and consequent are both present or both absent:

- p q
- not p not q

As a result, one endorses the MT inference ("not q" eliminates the first model and withdraws "not p"), as well as the AC and DA inferences ("q" eliminates the second model and keeps the first, and "not p" eliminates the first model and keeps the second), treating the conditional as if it were a biconditional. However, more sophisticated reasoners also consider the third possibility of a conditional:

- not p q

The additional information given by Rumain et al. (1983, Experiment 1) also draws the attention to this third possibility. Considering the third possibility, reasoners no longer make the AC and DA inferences. Furthermore, MT becomes more difficult because it is based on three explicit models.

Developmental studies (for a discussion, see Evans et al., 1993; see e.g. Barrouillet, Grosset, & Lecas, 2000, for a more recent study) support the level idea as far as the first two levels are concerned. At the first level, children reach the almost error-free performance of adults on MP, together with a good performance on MT. However, DA and AC are also frequently endorsed at this stage (Kodrof and Roberge,1975; Roberge, 1970). Subsequently, there is, up to adulthood, a decrease in performance on MT, as well as an increase in the capacity to resist the fallacies (Rumain et al., 1983, Experiment 1; Wildman and Fletcher, 1977). Hence, with increasing age, more and more individuals reach the second level. As even adults often fail in solving MT, no evidence for the existence of the third level of sophistication is found thus far. In the studies with a broad range of age groups (Rumain et al., 1983, Experiment 2; Barrouillet et al., 2000, Experiment 3), the results show a curvilinear pattern of accepting MT, DA and AC, first increasing and then decreasing. These results suggest that the first level is preceded by what could be called a zero-level, with an overall low acceptance (except for MP).

In assessing the level idea of Rumain et al. (1983) with developmental studies, there are two considerations

that question the appropriateness of these studies in addressing the level idea. First, the developmental pattern reveals an interpretation shift from a biconditional interpretation (by accepting the invited inferences) to a conditional interpretation (by no longer accepting the invited inferences). Since reasoning ability increases as a function of age, the developmental pattern supports the level idea. It is possible, however, that the relation between interpretation and propositional reasoning ability simply reflects an effect of age on both variables without a more intrinsic relation between both. Whether an intrinsic relation does exist, should be investigated by looking for it within age groups, holding age constant.

Second, in the developmental studies, no evidence for the existence of the third level of sophistication is found thus far, as even adults often fail in solving MT. It is possible however, that some of the adults do reach the third level, but that this is obscured in the results by the results of the majority that do not reach this level. Thus, the absence of evidence regarding the existence of the third level of sophistication is possibly due to a focus on the mean performance, without taking into account individual differences within a particular age group (e.g., adults).

Both considerations call for an individual differences approach, which allows for the assessment of the relation between interpretation and propositional reasoning ability within one age group, as well as for looking for the existence of a subgroup of sophisticated reasoners. In this article, we rely on the method of latent class analysis (Dayton, 1998; Goodman, 1974; Lazarsfeld & Henry, 1968) to capture individual differences.

We will conduct two different types of latent class analysis. The first type is an unrestricted latent class analysis on a first set of six problems, with an AC or DA component (among other components). We expect two classes, one with a biconditional interpretation accepting the fallacies, and another with a conditional interpretation coming to the conclusion that the problems are undecidable. The unrestricted latent class solution will be estimated empirically from the data without any restriction from the theory. If nevertheless the two expected classes appear, and if the classes are related to an external variable that is intended to measure the biconditional versus conditional interpretation using truth table tasks, then we have strong evidence for the presence of the two different interpretations in our sample.

The second type of latent class analysis is conducted on a second set of problems, consisting of 24 problems with a MP versus MT component (among other components). If the interpretation of the conditional has an impact on MT problems, and if the three-level hypothesis is valid, then we expect three classes: a first one with good performance on MT but a rather poor general propositional reasoning ability, a second one with a rather poor performance on MT but a better general propositional reasoning ability than the first class, and finally, a third class with again a good performance on MT and with the highest general propositional reasoning ability. We will use a confirmatory latent class analysis, with MP versus MT as a problem design factor that has a class specific effect on the probability of success besides the effect of other relevant design factors (Rijmen & De Boeck, 2001), and with a class specific intercept that can be interpreted as the general propositional reasoning ability of a particular latent class. In this way, the three-level hypothesis can be tested in a direct way. We expect of course that the first class corresponds with the biconditional class from the unrestricted latent class analysis on the first problem set consisting of problems with an AC or DA component, and that the second and third class correspond with the conditional class from the unrestricted analysis. Furthermore, we also expect a relation with the external measure of interpretation based on truth table tasks.[1]

To avoid a ceiling-effect (especially for problems with a MP component), all problems were constructed as complex reasoning problems, with additional components besides a DA or AC (first set) and MP or MT (second set) component. This was done by combining the four conditional inferences with other elementary inferences, such as disjunctive syllogism (DS; ''p or q'' and ''not p'' yields ''q''), conjunction (C; ''p'' and ''q'' yields ''p and q''), and disjunctive modus ponens (DMP; ''if p or q, then r'' and ''p'' yields ''r''). We refer to these complex problems with AC+ and DA+ (first set), and MP+ and MT+ (second set); (the ''+'' indicates that other inferences had to be made apart from AC, DA, MP, or MT). Part of the problems were already discussed in Rijmen and De Boeck (2001), addressing quite another question however.

Method

Participants

214 high school students between 16 and 19 years of age (fifth and sixth grade, Belgian school system) participated in the experiment.

[1]The authors opted for two different types of latent class analyses on the two problems sets instead of one joint analysis on all 30 problems together (the six problems of the first set and the 24 problems of the second set) for the following two reasons. First, for the set of six DA+ and AC+ problems, different interpretations of the conditional should lead to qualitatively different answers: conditional reasoners should answer 'undecidable', whereas biconditional reasoners should answer 'necessarily true'. For the set of 24 MP+ and MT+ problems, the correct answer is independent of the interpretation of the conditional. Interpretation is only supposed to affect the difficulty of those problems. The consequence is that the answers on the MP+ and MT+ problems can be dichotomized (correct/incorrect) without much loss of information, unlike the answers on the DA+ and AC+ problems, calling for a different modeling approach for the two sets of problems. Second, The two different types of analyses reflect two stages in modeling. We will first check *whether there are* latent classes corresponding with the interpretation of a conditional by conducting an *unrestricted* latent class analysis on the DA+ and AC+ problems. If the existence of different interpretations is revealed, we can be confident that it is appropriate to construct a *confirmatory* model for the set of MP+ and MT+ problems, based on the three-level hypothesis for conditional reasoning.

Material

Each participant received a booklet, with two pages of instructions, 10 pages of problems, and two pages with the truth table task. On each of the 10 pages of problems, three problems were presented, 30 problems in total. Under each problem, four response alternatives were presented:

- Necessarily true
- Necessarily not true
- Undecidable
- I don't know

It was explained in the instructions that participants should choose "necessarily true" if they believed that, given true premises, the conclusion *must be true*; "necessarily false" if they believed that the conclusion *must be false*; "undecidable" if they believed that, given the premises, the conclusion *could be true, but also could be false*, hence, that not enough information was given; and "I don't know" if they *could not work it out*. It was explained to the participants that "undecidable" was a genuine conclusion that could be correct for some of the problems, whereas "I don't know" means one either cannot choose among the other three or that one gives up solving the problem. The "I don't know" alternative was included in order to avoid guessing behaviour. Of the 30 problems, six were composite problems containing a AC or DA component (first set), and 24 were composite problems containing a MP or MT component (second set).

In the second task, on the last two pages, truth tables were elicited. For the truth tables, it was stressed in the instructions preceding the task that there were no wrong answers, and that the only purpose was to find out about the interpretation of sentence connectives. It was asked what the meaning was of some conditional propositions. For each proposition, the four possible combinations of the constituent elementary propositions (true-true, true-false, false-true, and false-false) were presented together with the question: "Is this combination possible given the proposition?" After the instructions, an example was given for a compound proposition with "and" (to avoid suggestive effects for the conditional propositions that followed the example):

- "Jan is in Paris and Frieda is in Barcelona"

In the example, the four possible combinations were given, together with the correct responses according to a conjunction:

• Jan is not in Paris; Frieda is not in Barcelona	no
• Jan is not in Paris; Frieda is in Barcelona	no
• Jan is in Paris; Frieda is not in Barcelona	no
• Jan is in Paris; Frieda is in Barcelona	yes

After this example, two conditional propositions were presented, one for each of two contexts, to be described later. The participants then had to circle "yes" or "no" for each of the four combinations following each proposition.

Design

Each participant had to evaluate the conclusion of 30 complex problems. The correct answer on six of the problems depended on the interpretation of the conditional (first set). For the other 24 (second set), the correct answer was independent of the interpretation of the conditional. All problems were constructed as a combination of elementary inferences (AC, DA, MP, MT, combined with DS, DMP, or MP and C).

The problems of the second set were constructed by combining premises as illustrated in Table 1, resulting in six problem types. A letter represents an elementary proposition, possibly containing a negation. It was never necessary to perform a double negation (e.g., for Problem Type 3, see Table 1, "q" was never a proposition with a negation).

Table 1 The six problem types

	MP	MT
MP + conjunction	1. if p, then q p if r, then s r – – – – – q and s	2. if p, then q p if r, then s not s – – – – – q and not r
DS	3. p or q if r, then not q r – – – – – p	4. p or q if q, then r not r – – – – – p
DMP	5. if p or q, then r if s, then p s – – – – – r	6. if p or not q, then r if q, then s not s – – – – – r

MP versus MT was the first factor in the design. For 12 reasoning problems, one inference was a MP (left column). For the other 12, this inference was a MT. The first problems are referred to as *MP + problems* (the "+" indicating that other inferences had to be made apart from MP), the latter are referred to as *MT + problems*.

The second factor of the design had three levels: *MP + C, DS, and DMP*. It refers to the inferences that are combined with MP or MT: modus ponens and conjunction, disjunctive syllogism, or disjunctive modus ponens. This factor is orthogonal with the first and makes a distinction between three sets of eight problems: For the first set of eight reasoning problems, the problems of *level MP + C*, (first row of Table 1), three inferences were to be made: the first one was a MP or a MT (see first factor), the second inference was a MP. This MP inference could be made independently from the MP or MT inference already discussed. The results of both conditional inferences must subsequently be combined with a conjunction (C), which was then the third inference. For the second set of eight problems, *the problems of level DS*, (second row of Table 1), there were only two inferences to be made: the first one was a MP or MT (see first factor), the second inference to be made was a disjunctive syllogism (DS). The categorical proposition of the DS was the result of the MP or MT inference from the first step. For the third set of eight problems, the problems of *level DMP*, (third row of Table 1), also two inferences were to be made: the first one was again a MP or MT, the second was a disjunctive modus ponens (DMP). The categorical proposition of the DMP was the result of the MP or MT inference from the first step.

A third factor, orthogonal to the first two, was the truth-value of the conclusion to be evaluated: *true or false*. For 12 problems the conclusion to be evaluated was true (true problems), for the other 12, the conclusion to be evaluated was false (false problems). For the latter, the conclusion to be evaluated contradicted the conclusion that followed from the premises. To avoid the presence of negations being a cue for the correct answer, some propositions of the premises contained a negation, though no problem required a double negation.

The fourth factor, manipulated orthogonally, was the *content of the problems*. Half of the problems were about people in cities or countries, for example "John is in Paris". The other half were about a functioning factory, for example, "The green light flashes". The orthogonal manipulation of these factors resulted into $2 \times 3 \times 2 \times 2 = 24$ problems.

A fifth factor, not orthogonally manipulated because this would render the experiment too long, was the *order of presentation of the premises*: grouped versus ungrouped. For each of both contents, the premises of the same inference were presented in a grouped way

Table 2 Class weights and class conditional probability estimates for the unrestricted model with three latent classes

Class weight Problem	Response category	Class 1	Class 2	Class 3
		.76	.20	.03
A	Necessarily true	.88	.21	.62
	Necessarily not true	.03	.00	.23
	Undecidable	.08	.79	.00
	I don't know	.01	.00	.14
B	Necessarily true	.89	.08	.43
	Necessarily not true	.04	.00	.00
	Undecidable	.08	.92	.14
	I don't know	.00	.00	.42
C	Necessarily true	.79	.40	.34
	Necessarily not true	.03	.08	.38
	Undecidable	.18	.52	.00
	I don't know	.00	.00	.28
D	Necessarily true	.48	.13	.00
	Necessarily not true	.20	.00	.71
	Undecidable	.30	.87	.00
	I don't know	.02	.00	.29
E	Necessarily true	.93	.31	.42
	Necessarily not true	.00	.00	.28
	Undecidable	.07	.69	.16
	I don't know	.00	.00	.14
F	Necessarily true	.90	.21	.69
	Necessarily not true	.02	.00	.15
	Undecidable	.07	.78	.16
	I don't know	.01	.02	.00

in half of the problems (as presented in Table 1), and separated by another premise in the other half of the problems (e.g., for Type 1: "p" became the first premise and "If p, then q" the third, with a premise of the other MP, "If r, then s", in between). For the content of a factory, these problems were of Type 2, 3 and 6, and for the content of people in cities or countries, these problems were of Type 1, 4 and 5.

The six problems for which the correct answer depended on the interpretation of the conditional were also constructed as a combination of elementary inferences, but now with a DA or an AC as the first inference instead of an MP or MT. For these problems, the correct answer was "necessarily true" under a biconditional interpretation, but "undecidable" under a conditional interpretation. Three of the problems contained an AC (denoted with B, E, and F in Table 2), three a DA (denoted with A, C, and D in Table 2). The other inferences to be made were equivalent to those with the MP and MT components.

Procedure

The experiment was conducted in groups of more or less 20 participants. At the beginning of the experiment, the instructions were briefly explained. Participants were asked not to guess, but to chose the response alternative "I don' t know" in case they either could not make one of the three evaluations: "necessarily true", "necessarily not true" or "undecidable", or in case they wanted to give up on the problem. It was stressed that they were not allowed to write down something while solving the problems. Participants had to complete the task within 50 minutes, and they all succeeded in doing so.

Results

Latent class analysis

In latent class analysis (Dayton, 1998; Goodman, 1974; Lazarsfeld & Henry, 1968), the association between the manifest variables, here the responses to the reasoning problems, is explained by a categorical latent variable that defines the latent classes. According to the model, the probability that a person n shows response pattern $\boldsymbol{y}_n$, given that he/she belongs to latent class t ($\zeta_n = t$; $t = 1,..., T$), can be written as a product of conditional probabilities (conditional on $\zeta_n = t$), one for each item,

$$p_{\mathbf{y}_n|\zeta_n=t} = \prod_{i=1}^{I} p_{y_{ni}|\zeta_n=t}, \tag{1}$$

where $p_{\mathbf{y}_n|\zeta_n=t}$ is the probability of response pattern $\boldsymbol{y}_n$, given latent class t, and $p_{y_{ni}|\zeta_n=t}$ is the probability for item i ($i = 1,..., I$) to be responded to with the response that is part of response pattern $\boldsymbol{y}_n$, given latent class t.

That is, the responses to the items are assumed to be conditionally independent, given class membership. The marginal probability that a randomly chosen participant will have response pattern $\boldsymbol{y}_n$ is

$$p_{\mathbf{y}_n} = \sum_{t=1}^{T} p_{\mathbf{y}_n|\zeta_n=t} \times p_t, \tag{2}$$

where p_t is the overall probability of belonging to class t, also called the class weight ($t = 1,..., T$).

After the parameters of the model are estimated (e.g., by using an expectation-maximization algorithm, Dempster, Laird, & Rubin, 1977), Bayes' theorem can be used to estimate the posterior probability that an individual belongs to class t, given that response pattern $\boldsymbol{y}_n$ has occurred,

$$p_{\zeta_n=t|\mathbf{y}_n} = \frac{p_{\mathbf{y}_n|\zeta_n=t} \times p_t}{p_{\mathbf{y}_n}}. \tag{3}$$

Each individual can then be allocated to the class with the maximum a posteriori probability.

We conducted two series of latent class analyses. In the first, the response patterns on the set of six AC+/DA+ problems were modeled with an unrestricted latent class model. That is, no constraints were placed on the class conditional probabilities $p_{y_{ni}|\zeta_n=t}$.

The second series of latent class analyses was conducted on the dichotomized (correct/ incorrect) response patterns of the second set of 24 MP+/MT+ problems. Here, a restricted latent class model was used. The restrictions on the class conditional probabilities $p_{y_{ni}|\zeta_n=t}$ were derived from the design of the complex reasoning problems, and from the three-level hypothesis (Rumain et al., 1983). The analyses on the two sets of problems are discussed next.

Unrestricted latent class analysis on the AC+/DA+ problems

The six AC+ and DA+ problems served as the manifest variables, each having four categories. We used the *l*EM (Vermunt, 1997) software packet for the unrestricted latent class analyses. For models with latent

variables, maximum likelihood estimates are obtained in *l*EM with an Expectation Maximization (EM) algorithm (Dempster et al., 1977).

Solutions were obtained for one (which is actually not a "real" latent class model, but a model in which independence of the responses is assumed), two, three and four latent classes. For each of the latter three, 10 analyses were carried out with different random starting values. Out of the 10 solutions, the solution with the highest loglikelihood was chosen. This was done in order to avoid ending up with a local maximum solution. The selected solutions always occurred more than once, so that we could be rather confident in having found the global maximum solutions.

Because of the large number of possible response patterns (4^6) in comparison with the observed number (214), the expected frequencies for most of the possible response patterns (based on the maximum likelihood estimates for the parameters of a particular model) are very small. In this case, the traditional goodness of fit statistics, such as the log-likelihood ratio G^2 or the Pearson X^2 test statistic, are of limited use to determine the number of classes because the statistics do not follow (asymptotically) the χ^2 reference distribution. Comparing the nested models (a two class model is nested within a three class model, a three class model within a four class model, and so on) with a difference G^2 statistic, which is less distorted for sparse data (Heinen, 1993), is not a solution either because a model with T latent classes is obtained by restricting parameters of the model with $T+1$ classes to their boundary values. In this case the regularity conditions under which the difference G^2 statistic follows a χ^2 reference distribution are no longer met. Furthermore, the difference G^2 statistic follows a χ^2 reference distribution only when the more general model is a true model, which we cannot determine, given our sparse data.

We addressed model selection by following a two-step procedure. First, a candidate model was selected based on the Bayesian information criterion (BIC) (e.g. Sclove, 1987):

$$BIC = -2 \times \text{loglikelihood} + (\text{number of estimated parameters}) \times \log(\text{sample size}) \quad (4)$$

The BIC is a parsimony index that weights the goodness of fit of the model (–2 × loglikelihood) against its parsimony (by adding a penalty term that increases with the number of estimated parameters). Lower values for the BIC indicate better solutions.

The BIC for the solutions with one, two, three, and four classes was respectively 2083.10, 1924.54, 1977.33, and 2055.83. Hence, the model with two classes was withdrawn as the candidate model.

In the second step, the fit of the candidate model was assessed using a parametric bootstrap procedure (Efron & Tibshirani, 1993) to approximate the reference distribution of two fit statistics: the Pearson X^2 and the Read-Cressie chi-squared (2/3) statistic. The two statistics perform quite well in a parametric bootstrap procedure to assess the fit of a latent class model when the data are sparse (von Davier, 1997). As a result, the solution with two classes was rejected: the Pearson X^2 amounted to 130392.63 (bootstrapped $p = .005$), and the Read-Cressie chi-squared (2/3) statistic amounted to 6911.17 (bootstrapped $p = .005$).

Based on the BIC, the second candidate model is the model with three classes. For the latter, the Pearson X^2 amounted to 2243.03 (bootstrapped $p = .08$), and the Read-Cressie chi-squared (2/3) statistic to 641.03 (bootstrapped $p = .065$). Hence, the model with three classes did provide an acceptable fit to the data.

A sufficient condition for identifiability of a latent class model is that the theoretical covariance matrix for the maximum likelihood estimates is positive definite (Dayton, 1998). For the solution with three classes, some of the eigenvalues of the observed information matrix, the inverse of which is an approximation of the theoretical covariance matrix of the maximum likelihood estimates, were zero or negative, indicating that the model was possibly not locally identified. However, non-positive eigenvalues also occur when parameters are too close to the boundary of the parameter space to check their identification (Vermunt, 1997). Inspecting the output revealed that the latter was probably the case, since the number of zero or negative eigenvalues corresponded with the number of class conditional probabilities with an estimate of zero (which is at the boundary of the parameter space [0, 1]). Furthermore the analyses in which the global optimum was found always resulted in the same class conditional probabilities. Hence, we can be confident that we found a unique solution.

The estimates of the class weights and the class conditional probabilities for the solution with three classes are given in Table 2.

The class weight of the first class amounted to .76, and, inspecting the conditional probabilities, it must be interpreted as the class of participants who adopted a biconditional interpretation, since the probabilities of responding with "necessarily true" is highest, and even very high, except in one case (Problem D). The second class, with a class weight of .20, corresponds with a conditional interpretation: these participants resisted to the invited inferences, since the probability of "undecidable" is clearly higher than that of other responses for all items. The third class is very small with a class weight of only .03. By consequence, there is only a limited amount of information on which the estimates of the class conditional probabilities are based, and one should be cautious in interpreting them. Insofar as the estimates of the class conditional probabilities can be trusted, it seems that the third class is a tiny class of random guessers, as several responses have a substantial probability, with a varying pattern across items. The third class is needed statistically, but is ignorable in size and has no substantive meaning. Therefore, we will continue with the first two classes only, and neglect the third class.

The interpretation of the two main classes can be corroborated by looking at the relation between allocated class membership and the responses on each of the truth table tasks. Thereto, the response patterns on the truth table tasks were classified for the two conditional propositions separately as corresponding with a "pure" conditional interpretation (always "yes", except for the combination " true antecedent – false consequent"), as corresponding with a biconditional interpretation (only "yes" for the combinations true antecedent – true consequent and false antecedent – false consequent) or as a rest category "other". For each of the two truth table tasks, participants belonging to the latter were omitted, as the truth table task was uninformative for those participants with respect to their interpretation of the conditional. Furthermore, each participant was assigned to the class with the maximum posterior class membership probability (using Equation 3). Participants allocated to the third class were also omitted. The class membership×response pattern classification tables are given in Table 3. For the truth table task with the content of a functioning factory, the Pearson X^2 is 55.69 ($df = 1$, $p < .001$), and for the truth table task with the content of people in cities or countries, the Pearson X^2 is 67.29 ($df = 1$, $p < .001$). Hence, class membership is related with the response pattern on each of the truth table tasks. The probabilities of belonging to the first class given a biconditional response pattern are respectively .88 for the content of a functioning factory and .94 the content of people in cities or countries. The probabilities of belonging to the second class given a conditional response pattern are respectively .8 and .63. These probabilities are always higher than the unconditional probabilities of belonging to the first or second class, respectively .78 and .22 for the content of a functioning factory, and .79 and .21 for the content of people in cities or countries. The small differences in unconditional probabilities between contents are due to the fact that the participants belonging to the rest category for the truth table task with the content of a functioning factory did not always belong to the rest category for the truth table task with the content of people in cities or countries, and vice versa, so that there was a slight difference in which participants were omitted in each of the two analyses.

Hence, based on the class conditional probabilities and the relation between allocated class membership and response pattern on the truth table tasks, we can conclude that the two classes correspond with a (large) class consisting of biconditonal reasoners, and a (smaller) class of conditional reasoners, as expected.

Restricted latent class analysis on the MP+/MT+ problems

A restricted latent class model was applied to the second set of problems, consisting of 24 dichotomized (correct/incorrect) MP+/MT+ problems. The restrictions on the class conditional probabilities were derived from a) the design of the complex reasoning problems, and b) the three level hypothesis of Rumain et al. (1983).

As explained in the "Design" section, the 24 MP+/MT+ problems were constructed as a combination of more elementary inferences: MP, MT, C, DS, and DMP, resulting in six problem types. Other independent variables were: whether the conclusion to be evaluated was true or false, the content of the problems, and the premise order. In a sequence of multiple regression analyses with as dependent variable the logit transformed proportions of correct answers for the 24 problems, Rijmen and De Boeck (2001) withheld four predictors: three variables effect coding for the six problem types, and a variable effect coding for whether the conclusion to be evaluated was true or false. Therefore, the logits of the class conditional probabilities of giving a correct answer on an item i ($p_{y_{ni}=1|\varsigma_n=t}$) were restricted to be a linear function of these four factors, according to the following logistic regression model:

$$\log\left(\frac{p_{y_{ni}=1|\varsigma_n=t}}{1-p_{y_{ni}=1|\varsigma_n=t}}\right) = \mathbf{x}'_{ni}\beta_t \qquad (5)$$

where $\mathbf{x}_{ni}$ is a five dimensional covariate vector item i, effect coding for the six problem types and whether the conclusion to be evaluated was true or false:

$x_{ni1} = 1$ for all i

$x_{ni2} = 1$ for the MP+ problems

$\quad -1$ for the MT+ problems

$x_{ni3} = 1$ for the problems with a MP + C

$\quad 0$ for the problems with DS

$\quad -1$ for the problems with a DMP

$x_{ni4} = 1$ for the problems with a DS

$\quad 0$ for the problems with a MP + C

$\quad -1$ for the problems with a DMP

$x_{ni5} = 1$ for the false problems

$\quad -1$ for the true problems

$\beta_t =$ the parameter vector for latent class t

$\quad$ (logistic regression weights)

Table 3 Classification tables of class membership for the two main classes of the unrestricted latent class model, and the biconditional versus conditional response patterns on the truth table tasks

Truth-table pattern	Functioning factory			People in cities or countries		
	Class 1	Class 2	Total	Class 1	Class 2	Total
Biconditional	126	18	144	119	7	126
Conditional	5	20	25	17	29	46
Total	131	38	169	136	36	172

β_t is defined as $\beta + \eta_t$, the sum of the vector of overall logistic regression weights (β), and the vector of class-specific deviations from the overall effects for latent class t (η_t). To identify the model, η_1 is restricted to $\mathbf{0}$. β and η_t, $t = 2, ..., T$, are the parameters actually estimated. A positive regression weight β_{tj} (j = 1, ..., 5) indicates that in class t items with a one on covariate x_j are the easiest, and problems with a minus one the most difficult, other things being equal. β_{t1}, the intercept for latent class t, can be interpreted as the overall propositional reasoning ability of class t, as the intercept parameter corresponds to the mean on a logistic scale of the solution probabilities of the 24 problems. The higher the intercept, the higher the general propositional reasoning ability of the members of class t.[2]

The latent classes of the restricted latent class model are expected to correspond with the different performance levels associated with the interpretation of conditionals. The different latent classes are expected to differ mainly with respect to the difficulty of MT, since a MT is predicted to be more difficult under a conditional interpretation. In addition, insofar as a conditional interpretation is indeed associated with a higher general propositional reasoning ability, the latent classes are expected to differ in this respect also. The difficulties of the other elementary inferences (MP, C, DS, DMP) are not expected to differ across classes. These expectation can be modeled by allowing a class-specific weight for x_1, the intercept (reasoning ability), and x_2 (the relative difficulty of MT in comparison to MP), and restricting the other logistic regression weights to be the same across classes.

Finally, we allowed for differences in reasoning ability within latent classes also. The latter was incorporated by assuming that the intercept was random within each class. More specifically, we assumed that the intercept followed a normal distribution over the population of participants belonging to the same class.

Formally, given latent class t and reasoning ability θ_{nt} (within class t), the class conditional probabilities are modeled as

$$\log\left(\frac{p_{y_{ni}=1|\varsigma_n=t,\theta_{nt}}}{1-p_{y_{ni}=1|\varsigma_n=t,\theta_{nt}}}\right)=\mathbf{x}'_{ni}\beta_t+\theta_{nt}, \tag{6}$$

or equivalently,

$$p_{y_{ni}=1|\varsigma_n=t,\theta_{nt}}=\frac{\exp\left(\mathbf{x}'_{ni}\beta_t+\theta_{nt}\right)}{1+\exp(\mathbf{x}'_{ni}\beta_t+\theta_{nt})}, \tag{7}$$

where $\boldsymbol{x}_{ni}$ and β_t are as defined in Equation 5, but with the restriction that $\beta_{1j} = ... = \beta_{Tj}$, or alternatively, $\eta_{1j} = ... = \eta_{Tj} = 0$, for $j = 3, 4, 5$ (all effects but the intercept and the effect of MP versus MT); and

θ_{nt} are independent and identically distributed as $N(0,\sigma^2)$.

The variances of θ_{nt} were assumed equal across classes in order to keep the number of parameters as low as possible.

Within a latent class t, the marginal probability of response pattern $\boldsymbol{y}_n$ is then

$$p_{\mathbf{y}_n|\varsigma_n=t}=\int_{\theta_{nt}}\left(\prod_{i=1}^{I}\frac{\exp\left[y_{ni}\left(\mathbf{x}'_{ni}\beta_t+\theta_{nt}\right)\right]}{1+\exp(\mathbf{x}'_{ni}\beta_t+\theta_{nt})}\right)N\left(\theta_{nt}|0,\sigma^2\right)d\theta_{nt}. \tag{8}$$

This is the marginal formulation of a logistic regression model with a random intercept. For logistic regression models with random effects, see for example Hedeker and Gibbons (1994), and Zeger and Karim (1991). These models are equivalent with a broad range of item response theory models (Rijmen, Tuerlinckx, & De Boeck; 2001).

Marginalizing over latent classes, the response probability of response pattern $\boldsymbol{y}_n$ is:

$$p_{\mathbf{y}_n}=\sum_{t=1}^{T}p_t\left\{\int_{\theta_{nt}}\left(\prod_{i=1}^{I}\frac{\exp\left[y_{ni}\left(\mathbf{x}'_{ni}\beta_t+\theta_{nt}\right)\right]}{1+\exp(\mathbf{x}'_{ni}\beta_t+\theta_{nt})}\right)N\left(\theta_{nt}|0,\sigma^2\right)d\theta_{nt}\right\}. \tag{9}$$

Discrete mixture models, where each mixture is an IRT model (logistic regression model with random intercept) have been described by Mislevy and Verhelst (1990), Mislevy and Wilson (1996), Rost (1990), Yamamoto (1987), and Wilson (1989).

The maximum likelihood estimates of the model described in Equation 9 were obtained using an EM algorithm, which is described in the Appendix.

To find the appropriate model, the same procedure was followed as in the unrestricted latent class analyses. Solutions were obtained for one, two, three and four latent classes. For each of the latter three, 10 analyses were carried out with different starting values. Out of the 10 solutions, the solution with the highest loglikelihood was chosen. For the two and three classes analyses, always the same solution was found. For the four classes analyses, the solution with the highest loglikelihood occurred eight times out of 10. Hence, we can be confident in having found the global maximum solutions.

The BIC for the solutions with one, two, three, and four classes were respectively 5316.8, 5270.5, 5255.8, and 5268.4. Hence, the three class model was selected as the candidate model.

In the second step, the fit of the candidate model again was assessed by a parametric bootstrap procedure. For discrete mixtures, where each component is an IRT model, von Davier (1997) also recommends using the Pearson X^2 and the Read-Cressie chi-squared (2/3) statistics. the Pearson X^2 amounted to 26777214.09 (bootstrapped $p = .095$), and the Read-Cressie chi-squared (2/3) 376765.61 (bootstrapped $p = .205$). Hence the

[2]We would like to thank an anonymous reviewer for suggesting effect coding for the independent variables of the logistic regression, so that the intercept corresponds to the mean on a logistic scale of the solution probabilities.

Table 4 Parameter estimates, standard errors and approximate .05 confidence intervals (parameter estimates +-1.96 times the standard errors) for the restricted latent class model

	Parameter estimate	Standard error	.05 confidence interval
p_1 (class weight of class 1)	.70	.05	(.60; .81)
p_2 (class weight of class 2)	.26	.05	(.15; .36)
p_3 (class weight of class 3)	.04	–	–
β_1 (overall intercept)	.82	.07	(.67; .97)
η_{11} (deviation from overall intercept for class 1)	0	–	–
η_{21} (deviation from overall intercept for class 2)	1.92	.21	(1.51; 2.33)
η_{31} (deviation from overall intercept for class 3)	–.40	.32	(–1.02; 22)
β_2 (overall effect of MP vs. MT)	.37	.05	(.28; .46)
η_{12} (deviation from overall effect of MP vs. MT for class 1)	0	–	–
η_{22} (deviation from overall effect of MP vs. MT for class 2)	.36	.18	(.00; .72)
η_{32} (deviation from overall effect of MP vs. MT for class 3)	2.20	.26	(1.69; 2.71)
β_3 (MP + C = 1; DS = 0; DMP = −1)	.86	.05	(.76; .96)
β_4 (DS = 1; MP + C = 0; DMP = −1)	–.84	.04	(–.93; –.76)
β_5 (false vs. true)	–.28	.04	(–.35; –.21)
σ^2 (variance random intercept)	.11	.07	(0; .25)

Table 5 Class weights and marginal class conditional probabilities of giving a correct answer for the six problem types for the restricted model with three latent classes

Class weight	Class 1		Class 2		Class 3	
	.70		.26		.04	
	True problems	False problems	True problems	False problems	True problems	False problems
Problem type 1	.91	.85	.99	.98	.98	.97
Problem type 2	.83	.73	.96	.93	.27	.18
Problem type 3	.65	.52	.95	.91	.92	.86
Problem type 4	.47	.34	.80	.70	.07	.04
Problem type 5	.81	.70	.98	.96	.96	.93
Problem type 6	.67	.54	.90	.84	.14	.08

model with three latent classes provided an acceptable fit.

The observed information matrix was positive definite, so that we can be confident that the model was locally identified.

The parameter estimates, standard errors and approximate .05 confidence intervals (parameter estimates +-1.96 times the standard errors) are given in Table 4.

As in the unrestricted latent class analysis, there is a large class containing the majority of the participants (Class 1, $p_1 = .70$), a smaller one (Class 2, $p_2 = .26$), and a very tiny third class (Class 3, $p_3 = .04$). The logistic regression weights have to be interpreted as follows:

1. Reasoning ability is highest for the second class ($\beta_{21} = .82 + 1.92 = 2.74$), intermediate for the first class ($\beta_{11} = .82$) and lowest for the third class ($\beta_{31} = .82 - .40 = .42$). Looking at the confidence intervals, we can conclude that the difference between the first and the second class reaches significance ($\eta_{21} = 1.92$ differs significantly from zero), but the difference between the first and the third class does not ($\eta_{31} = -.40$ is not significantly different from zero).
2. The relative difficulty of MT in comparison with MP is highest for the third class ($\beta_{32} = .37 + 2.20. = 2.57$), intermediate for the second class ($\beta_{22} = .37 + .36 = .73$), and lowest for the first class ($\beta_{12} = .37$). Both the differences between the first class and the second class ($\eta_{22} = .36$), and between the first and the third class ($\eta_{32} = 2.20$) are significant, though only marginally for the former.
3. Problems with a MP + C are easier than problems with a DMP, which in turn are easier than problems with a DS ($\beta_3 = .86$ and $\beta_4 = -.84$, both significantly different from zero).
4. False problems are more difficult than true problems ($\beta_5 = -.28$, significantly different from zero).
5. The variability in reasoning ability within classes is rather small ($\sigma^2 = .11$).

The marginal (integrated over θ_{nt}) class conditional probabilities of giving a correct answer on the six problem types are given in Table 5, for the true and false problems separately.

Based on the parameter estimates and the marginal class conditional probabilities, we can interpret the first and second class as corresponding with respectively the level of unsophisticated and the level of intermediate reasoners: The members of the second class have relatively more difficulties with MT (in comparison with MP) than the members of the first class, but also a higher general propositional reasoning ability. The higher general propositional reasoning ability of the second class outweighs the greater difficulty with MT, so that the probabilities of giving a correct answer on the MT+ problems is still always higher for the second class

Table 6 Classification tables of class membership for the two main classes of the restricted latent class model, and the biconditional versus conditional response patterns on the truth table tasks

Truth-table pattern	Functioning factory			People in cities or countries		
	Class 1	Class 2	Total	Class 1	Class 2	Total
Biconditional	108	36	144	96	32	128
Conditional	11	13	24	24	17	41
Total	119	49	168	120	49	169

than for the first class. Hence, the MT+ problems are only relatively, but not absolutely, more difficult for reasoners of an intermediate level than for reasoners of an unsophisticated level. The third class does not correspond with the level of sophisticated reasoners. However, it is a very tiny class. Given the ignorable size of the third class, we will continue with the first two classes only, as we did in the unrestricted latent class analysis.

The interpretation of the two main classes was corroborated by the relation between allocated class membership and the responses on each of the truth table tasks. Again, each participant was assigned to the class with the maximum posterior class membership probability (using Equation 3); and participants allocated to the third class or with an uninformative response pattern on the truth table task were omitted. The membershipresponse pattern classification tables are given in Table 6. For the truth table task with the content of a functioning factory, the Pearson X^2 is 8.47 ($df = 1$, $p = .002$), and for the truth table task with the content of people in cities or countries, the Pearson X^2 is 4.09 ($df = 1$, $p = .02$). Hence, class membership is related with the response pattern on each of the truth table tasks. The probabilities of belonging to the first class given a biconditional response pattern are .75 for both contents. The probabilities of belonging to the second class given a conditional response pattern are .54 for the content of a functioning factory and .41 for the content of people in cities or countries. These probabilities are always higher than the unconditional probabilities of belonging to the first or second class, respectively .71 and .29 for both contents. Though the expected relations between class membership and response pattern on the truth table tasks is less strong than was the case for the unrestricted latent class analysis, they turn up consistently.

Finally, the two main classes from the unrestricted analysis are related to the two main classes of the restricted analysis, omitting the participants allocated to the third class in either of the two latent class models. The classification is given in Table 7. The Pearson X^2 is 7.01 ($df = 1$, $p = .008$). The probability of belonging to the first class in the restricted (unrestricted) model given that one belongs to the first class in the unrestricted (restricted) model is .77 (.86), and the probability of belonging to the second class in the restricted (unrestricted) model given that one belongs to the second class in the unrestricted (restricted) model is .44 (.30). These probabilities are higher than the unconditional probabilities of belonging to the first or second class of the restricted (unrestricted) model, respectively .73 (.82) and .27 (.18). Hence, as expected, the first class of the unrestricted model is associated with the first class of the restricted model, and the second class of the unrestricted model is associated with the second class of the restricted model.

Table 7 Classification tables of class membership for the two main classes of the restricted latent class model, and the two main classes of the unrestricted latent class model

Unrestricted latent class model	Restricted latent class model		
	Class 1	Class 2	Total
Class 1	125	37	162
Class 2	20	16	36
Total	145	53	198

General discussion

A main finding in the area of conditional reasoning is that people often endorse DA and AC, though they are not valid in classical logic. This can be explained if one accepts that people treat the conditional not as a pure conditional, but as a biconditional, either by accepting the invited inferences (rule theories), or by neglecting the "false antecedent"-"true consequent" model of the conditional (mental model theory). Indeed, blocking the biconditional interpretation, Rumain et al. (1983) found that people were less inclined to accept DA and AC. However, also MT was endorsed less. Rumain et al. (1983) explained their results by distinguishing between three levels in the sophistication of conditional reasoning:

1. Unsophisticated reasoners solve a MT by accepting the invited inferences.
2. Reasoners of an intermediate level can resist the invited inferences, but cannot find the more difficult reductio ad absurdum line of reasoning to solve a MT.
3. Sophisticated reasoners do not draw the invited inferences either, but do master the reductio strategy and hence can solve a MT.

We distinguish three aspects within this explanation:

1. A conditional interpretation hinders the performance on MT in comparison with a biconditional treatment of the conditional.
2. A conditional interpretation is associated with a higher general propositional reasoning ability.
3. There are three levels of performance.

The first aspect is corroborated in an indirect way, by the experiment of Rumain et al. (1983) in which the biconditional interpretation is blocked, and by the developmental pattern in which a decrease in AC and DA acceptance is paralleled by a decrease in the endorsement of MT, as well as in a direct way (Johnson-Laird et al., 1992).

The second aspect is corroborated by the developmental pattern: the increase in reasoning ability as a function of age is paralleled by a shift from a biconditional towards a conditional interpretation. However, as explained in the introduction, developmental studies fail in providing evidence for an intrinsic relation between the interpretation of the conditional and general propositional reasoning ability.

With respect to the third aspect, developmental studies corroborate the existence of the first two levels. No evidence is found for the existence of a subgroup of sophisticated reasoners hitherto.

In this study, we investigated the three aspects using latent class models. Especially for the latter two, the individual differences approach of the latent class analysis allows for a more fine-tuned investigation. Using only one age group, it can be tested whether there is an intrinsic relation between the interpretation of the conditional that people spontaneously adopt and general propositional reasoning ability. Furthermore, by looking for the existence and the characteristics of several subgroups instead of focusing on the mean performance, a subgroup of sophisticated reasoners should be found more readily if present.

The unrestricted latent class model on the first problem set revealed the existence of two main subgroups. Based on the estimated class conditional probabilities, the subgroups could be interpreted in terms of the interpretation of the conditional spontaneously adopted: a large subgroup with a biconditional interpretation, and a smaller one with a conditional interpretation.

Given the results of the unrestricted latent class model, revealing different interpretations of the conditional, it was worthwhile to perform a restricted latent class analysis on the second set of problems, consisting of 24 problems with a MP or MT component. The restrictions were derived from the three level hypothesis of Rumain et al. (1983) about conditional reasoning. Again, two substantial classes were found. The largest class could be interpreted as consisting of unsophisticated reasoners; the second class, considerably smaller, as consisting of reasoners of an intermediate level. Hence, the first two aspects of the three level hypothesis were corroborated: a hindering effect of a conditional interpretation on MT, and the association of a conditional interpretation with a higher general propositional reasoning ability. Especially the latter finding is important, evidence for an intrinsic relation between interpretation and general propositional reasoning ability was lacking hitherto. No evidence was found for the existence of a subgroup of sophisticated reasoners. However, our participants were adolescents. It is possible that a subgroup of sophisticated reasoners will show up in a sample of adults.

The interpretation of the latent classes for both the unrestricted and the restricted latent class model was corroborated by comparing the classifications of the participants based on the posterior probabilities of belonging to a class with each other and with a classification based on the truth table tasks. The biconditional reasoners of the unrestricted latent class model were associated with the unsophisticated reasoners of the restricted model, and with a biconditional response pattern on the truth table tasks. The conditional reasoners of the unrestricted model were associated with the reasoners of an intermediate level of the restricted model, and with a conditional response pattern on the truth table tasks.

That a conditional interpretation has a detrimental effect on the performance on MT, but is also related with a higher general propositional reasoning ability, is an interesting example of how being smart might make it more difficult to solve a problem, or how a better understanding makes things more complex and therefore more difficult.

Appendix

EM algorithm for constrained latent class analysis

Our use of the EM algorithm for the discrete mixture model of Equation 9 resembles the use of the EM algorithm as proposed by Mislevy and Verhelst (1990), and Mislevy and Wilson (1996) for these kind of models.

The marginal likelihood function for the model of Equation 9 is

$$L(\delta, \mathbf{p}, \sigma^2|\mathbf{Y}) = \prod_{n=1}^{N}\sum_{t=1}^{T} p_t \left\{ \int_{\theta_{nt}} \left(\prod_{i=1}^{I} \frac{\exp\left[y_{ni}\left(\mathbf{x}'_{nit}\delta + \theta_{nt}\right)\right]}{1+\exp(\mathbf{x}'_{nit}\delta + \theta_{nt})} \right) \times N\left(\theta_{nt}|0,\sigma^2\right) d\theta_{nt} \right\}, \qquad (10)$$

where δ is a parameter vector containing both the class specific logistic regression weights and the logistic regression weights constant across classes; $\mathbf{x}'_{nit}$ is a class specific covariate vector, defined such that $\mathbf{x}'_{nit}\delta = \mathbf{x}'_{ni}\beta_t$. $\boldsymbol{p} = (p_1, \ldots, p_T)$ with $p_T = 1 - \sum_{t=1}^{T-1} p_t$; and $\boldsymbol{Y}$ is the $N \times \mathbf{I}$ datamatrix.

Using $\mathbf{x}'_{nit}$ and δ instead of $\mathbf{x}'_{ni}$ and β_t is done for computational convenience. It allows to treat all logistic

regression weights as logistic regression weights that are constant across classes. If a logistic regression weight β_{tj} is class specific for a particular class t, this can be taken into account by setting $x_{nit'j}$ to zero for all $t' \neq t$.

The idea behind the EM-algorithm is that maximizing (10) with respect to the parameters would be easier if the ς_n and θ_{nt} were known. In this case, the complete data (consisting of the observed data $\boldsymbol{Y}$ and the "missing data" ς_n and θ_{nt}) likelihood L_c reduces to:

$$L_c\left(\delta,p,\sigma^2|\mathbf{Y},\sigma,\theta_\varsigma\right)=\prod_{n=1}^{N}p_t\left(\prod_{i=1}^{I}\frac{\exp\left[y_{ni}\left(\mathbf{x}'_{nit}\delta+\theta_{nt}\right)\right]}{1+\exp\left(\mathbf{x}'_{nit}\delta+\theta_{nt}\right)}\right)N\left(\theta_{nt}|0,\sigma^2\right) \tag{11}$$

Where: ς and θ_ς are $N\times1$ vectors, containing for each person n the class t he/she belongs to and the ability θ_{nt} of that person within the class t he/she belongs to.

The missing data being unknown (by definition), the EM algorithm iterates as follows: at iteration $k+1$,

1. Compute the expected complete data loglikelihood, given the observed data and provisional parameter estimates obtained in iteration k.
2. Maximize the expected complete data loglikelihood with respect to the parameters of interest, resulting in new provisional parameter estimates for iteration $k+1$.

Iterate until some convergence criterion is met. For the first iteration, initial parameter estimates can be chosen at random or based on some preliminary analysis. Dempster et al. (1977) show that, under regularity conditions, the likelihood is not decreased after an iteration.

The expected complete data loglikelihood, given the observed data and the provisional parameter estimates is:

$$\begin{aligned}Q\left(\alpha|\alpha^{(k)}\right)&=E_{\alpha^{(k)}}[\log L_c(\alpha|\mathbf{Y},\sigma,\theta_\sigma)|\mathbf{Y}]\\&=\sum_{n=1}^{N}\sum_{t=1}^{T}p\left(\varsigma_n=t|\mathbf{y}_n,\alpha^{(k)}\right)\\&\quad\times\int_{\theta_{nt}}\log g(\mathbf{y}_n,\varsigma_n=t,\theta_{nt}|\alpha)f\\&\quad\times\left(\theta_{nt}|\varsigma_n=t,\mathbf{y}_n,\alpha^{(k)}\right)d\theta_{nt}\end{aligned} \tag{12}$$

where

α is the vector of all parameters (δ, $\boldsymbol{p}$, σ^2),
$\alpha^{(k)}$ is the vector of provisional parameter estimates obtained in iteration k, and
$E_{\alpha^{(k)}}$ denotes the expectation using parameter vector $\alpha^{(k)}$.

Differentiating $Q(\alpha|\alpha^{(k)})$ with respect to the parameters α results in the following score equations:

$$\frac{\partial Q\left(\alpha|\alpha^{(k)}\right)}{\partial p_t}=\sum_{n=1}^{N}\left[p\left(\varsigma_n=t|\mathbf{y}_n,\alpha^{(k)}\right)/p_t-p\left(\varsigma_n=T|\mathbf{y}_n,\alpha^{(k)}\right)/(p_T)\right] \tag{13}$$

for $t = 1,..., T-1$

$$\begin{aligned}\frac{\partial Q\left(\alpha|\alpha^{(k)}\right)}{\partial\sigma^2}&=\sum_{n=1}^{N}\sum_{t=1}^{T}p\left(\varsigma_n=t|\mathbf{y}_n,\alpha^{(k)}\right)\\&\quad\times\int_{\theta_{nt}}\left[-1/2\sigma^2+\theta_{nt}^2/2\sigma^4\right]f\left(\theta_{nt}|\varsigma_n=t,\mathbf{y}_n,\alpha^{(k)}\right)d\theta_{nt},\end{aligned} \tag{14}$$

and

$$\begin{aligned}\frac{\partial Q\left(\alpha|\alpha^{(k)}\right)}{\partial\delta_j}&=\sum_{n=1}^{N}\sum_{t=1}^{T}p\left(\varsigma_n=t|\mathbf{y}_n,\alpha^{(k)}\right)\\&\quad\times\left\{\sum_{i=1}^{I}y_{ni}x_{nitj}-\int_{\theta_{nt}}\sum_{i=1}^{I}\frac{\exp\left(\theta_{nt}+\mathbf{x}'_{nit}\delta\right)x_{nitj}}{1+\exp\left(\theta_{nt}+\mathbf{x}'_{nit}\delta\right)}f.\right.\\&\quad\left.\times\left(\theta_{nt}|\varsigma_n=t,\mathbf{y}_n,\alpha^{(k)}\right)d\theta_{nt}\right\}\end{aligned} \tag{15}$$

By equating the partial derivatives to zero, closed form solutions can be found for $p_1,...p_{T-1}$, and σ^2 that maximize $Q(\alpha|\alpha^{(k)})$, serving as provisional parameter estimates for the E-step of iteration $k+2$:

$$p_t^{(k+1)}=\frac{\sum_{n=1}^{N}p\left(\varsigma_n=t|\mathbf{y}_n,\alpha^{(k)}\right)}{N}\quad\text{for } t=1,...,T-1, \tag{16}$$

and

$$\sigma^{2(k+1)}=\frac{\sum_{n=1}^{N}\sum_{t=1}^{T}p\left(\varsigma_n=t|\mathbf{y}_n,\alpha^{(k)}\right)\int_{\theta_{nt}}\theta_{nt}^2f\left(\theta_{nt}|\varsigma_n=t,\mathbf{y}_n,\alpha^{(k)}\right)d\theta_{nt}}{N}. \tag{17}$$

For the logistic regression weights δ_j, no closed form solution exists. Instead, the parameter estimates are updated using one Newton-Raphson iteration. Hence actually, the algorithm we implemented is a Generalised EM algorithm (Dempster et al., 1977). The second derivatives needed for the Newton-Raphson algorithm are

$$\begin{aligned}\frac{\partial^2Q\left(\alpha|\alpha^{(k)}\right)}{\partial\delta_j\partial\delta_{j'}}&=-\sum_{n=1}^{N}\sum_{t=1}^{T}p\left(\varsigma_n=t|\mathbf{y}_n,\alpha^{(k)}\right)\\&\quad\times\int_{\theta_{nt}}\left\{\sum_{i=1}^{I}\frac{\exp\left(\theta_{nt}+\mathbf{x}'_{nit}\delta\right)x_{nitj}x_{nitj'}}{\left[1+\exp\left(\theta_{nt}+\mathbf{x}'_{nit}\delta\right)\right]^2}\right\}\\&\quad\times f\left(\theta_{nt}|\varsigma_n=t,\mathbf{y}_n,\alpha^{(k)}\right)d\theta_{nt}.\end{aligned} \tag{18}$$

The posterior densities $f(\theta_{nt}|\varsigma_n = t, y_n, \alpha^{(k)})$and the posterior probabilities $p(\varsigma_n = t|y_n, \alpha^{(k)})$ are computed using Bayes' theorem:

$$f\left(\theta_{nt}|\varsigma_n=t,\mathbf{y}_n,\alpha^{(k)}\right)=\frac{p(\mathbf{y}_n|\varsigma_n=t,\theta_{nt},\delta)N\left(\theta_{nt}|0,\sigma^2\right)}{\int\limits_{\theta_{nt}}p(\mathbf{y}_n|\varsigma_n=t,\theta_{nt},\delta)N(\theta_{nt}|0,\sigma^2)d\theta_{nt}} \tag{19}$$

$$p\left(\varsigma_n=t|\mathbf{y}_n,\alpha^{(k)}\right)=\frac{p_t\int\limits_{\theta_{nt}}p(\mathbf{y}_n|\varsigma_n=t,\theta_{nt},\delta)N\left(\theta_{nt}|0,\sigma^2\right)d\theta_{nt}}{\sum\limits_{t=1}^{T}p_t\int\limits_{\theta_{nt}}p(\mathbf{y}_n|\varsigma_n=t,\theta_{nt},\delta)N(\theta_{nt}|0,\sigma^2)d\theta_{nt}} \tag{20}$$

The integrals over θ_{nt} are approximated with Gaussian quadrature.

After convergence, the observed information matrix for the maximum likelihood estimates $\hat{\alpha}$ can be approximated by

$$I(\hat{\alpha}|\mathbf{Y})\approx\sum_{n=1}^{N}\mathbf{s}(\mathbf{y}_n|\hat{\alpha})\mathbf{s}'(\mathbf{y}_n|\hat{\alpha}) \tag{21}$$

where $s(\mathbf{y}_n|\hat{\alpha})=\partial\log L_n\left(\delta,\mathbf{p},\sigma^2|\mathbf{y}_n\right)/\partial\alpha$, the score function based on the single observation $\boldsymbol{y}_n$. The latter can be computed using the complete data loglikelihood as follows (McLachlan & Krishnan, 1997):

$$\mathbf{s}(\mathbf{y}_n|\alpha)=E_\alpha[\partial\log L_{cn}(\alpha|\mathbf{y}_n,\varsigma_n,\theta_{nt})/\partial\alpha|\mathbf{y}_n] \tag{22}$$

where $L_{cn}(\alpha|y_n, \varsigma_n, \theta_{nt})$is the complete data loglikelihood formed from the single observation $\boldsymbol{y}_n$.

References

Barrouillet, P., Grosset, N., & Lecas, J.F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition, 75,* 237–266.

Braine, M.D.S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85*, 1–21.

Braine, M.D.S., Reiser, B.J., & Rumain, B. (1984). Some empirical justification for a theory of natural propositional logic. In G. Bower (Ed.), *The psychology of learning and motivation*. (Vol. 18, pp. 313–371). New York: Academic Press.

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.

Dayton, C.M. (1998). *Latent class scaling analysis*. Thousand Oaks CA: Sage.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM-algorithm (with discussion). *Journal of the Royal Statistical Society, B, 39*, 1–38.

Efron, B., Tibshirani, R.J. (1993). *An introduction to the bootstrap*. New York: Chaman & Hall.

Evans, J. St. B.T., Newstead, S.E., & Byrne, R.M.J. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Erlbaum.

Geis, M.C., & Zwicky, A.M. (1971). On invited inferences. *Linguistic Inquiry, 2*, 561–566.

Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*, 215–231.

Hedeker, D., & Gibbons, R.D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics, 50*, 933–944.

Heinen, T. (1993). *Discrete latent variable models*. Series on work and organization. Tilburg, The Netherlands: Tilburg University Press.

Johnson-Laird, P.N., & Byrne, R.M.J. (199l). *Deduction*. Hillsdale, NJ: Erlbaum.

Johnson-Laird, P.N., Byrne, R.M.J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99*, 418–439.

Kodroff, J.K., & Roberge, J.J. (1975). Developmental analysis of the conditional reasoning abilities of primary grade children. *Developmental Psychology, 13*, 342–353.

Lazarsfeld, P.F., and Henry, N.W. (1968). *Latent structure analysis*. Boston: Houghton Mill.

McLachlan, G., & Krishnan, T. (1997). *The EM algorithm and extensions*. New York: Wiley.

Mislevy, R.J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika, 55,* 195–215.

Mislevy, R.J., & Wilson, M. R. (1996). Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrika, 61,* 41–71.

Osherson, D.N. (1975). *Logical abilities in children: Vol 3. Reasoning in adolescence: Deductive inference*. Hillsdale, NJ: Erlbaum.

Rijmen, F., & De Boeck, P (2001). Propositional reasoning: The differential contribution of "rules" to the difficulty of complex reasoning problems. *Memory & Cognition, 29,* 165–175.

Rijmen, F., Tuerlinckx, F., & De Boeck, P. A nonlinear mixed model framework for IRT models. Submitted for publication.

Rips, L.J. (1983). Cognitive processes in propositional reasoning. *Psychological Review, 90*, 38–71.

Rips, L.J. (1994). *The psychology of proof*. Cambridge, IM: Routledge.

Roberge, J.J. (1970). A study of children s ability to reason with basic principles of deductive reasoning. *American Educational Research Journal, 7*, 583–596.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14,* 271–282.

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 Psychology*, *19*, 471–481.

Sclove, S.L. (1987). Application of model selection criteria to some problems in multivariate analysis. *Psychometrika, 52*, 333–343.

Vermunt, J.K. (1997). *lEM: A general program for the analysis of categorical data*. Tilburg, The Netherlands, Tilburg University.

Von Davier, M. (1997). Bootstrapping goodness-of-fit statistics for sparse categorical data: Results of a Monte Carlo study. *Methods of* Psychological *Research Online, 2,* 29–48.

Wildman,T.M., & Fletcher, H.J. (1977). Developmental increases and decreases in solutions of conditional syllogism problems. *Developmental Psychology, 13*, 630–636.

Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin, 105*, 276–289.

Yamamoto, K. (1987). *A hybrid model for item responses*. Unpublished dissertation, University of Illinois.

Zeger, S.L., & Karim, M.R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association, 86*, 79–86.

Copyright © 2003 EBSCO Publishing