A Mixed Rasch Model of Dual-Process Conditional Reasoning

Jean-François Bonnefon
Université de Toulouse

Michael Eid
Free University of Berlin

Stéphane Vautier and Saïd Jmel
Université de Toulouse

A fine-grained dual-process approach to conditional reasoning is advocated: Responses to conditional syllogisms are reached through the operation of either one of two systems, each of which can rely on two different mechanisms. System1 relies either on pragmatic implicatures or on the retrieval of information from semantic memory; System2 operates first through inhibition of System1, then (but not always) through activation of analytical processes. It follows that reasoners will fall into one of four groups of increasing reasoning ability, each group being uniquely characterized by (a) the modal pattern of individual answers to blocks of affirming the consequent, denying the antecedent, and modus tollens syllogisms featuring the same conditional; and (b) the average rate of determinate answers to AC, DA, and MT. This account receives indirect support from the extant literature, and direct support from a mixed Rasch model of responses given to 18 syllogisms by 486 adult reasoners.

The capacity to solve conditional syllogisms is considered the cornerstone of both deductive ability and hypothetical thinking (Evans & Over, 2004), and psychological studies abound on how and how well reasoners answer the four standard problems called modus ponens (if *p* then *q*, *p*), modus tollens (if *p* then *q*, *not*−*q*), affirming the consequent (if *p* then *q*, *q*), and denying the antecedent (if *p* then *q*, *not*−*p*). Henceforth, these four problems will be abbreviated MP, MT, AC, and DA, respectively. Considering that almost everyone solves MP correctly most of the time by drawing the conclusion '*q*' we will focus on responses to AC, DA, and MT; from the perspective of standard deductive logic, the determinate conclusion '*not*−*p*' is correct for MT, and the undeterminate response 'one cannot draw any conclusion' is correct for AC and DA.[1]

In this article, we will elaborate a fine-grained dual-process account of how reasoners solve these three problems, define the psychometric model that reflects this account, and test that model against answers given by an unusually large sample of adult reasoners varying widely in age, occupation, and education. We will in particular argue that our approach to statistical modelling, that is, the use of mixed Rasch models, is uniquely suited to evaluate a complex dual-process account of reasoning because it allows for the simultaneous modelling of qualitative and quantitative differences between individuals.

Dual-Process Conditional Reasoning

Dual-process theories of reasoning (Evans & Over, 1996; Sloman, 1996; Stanovich, 1999) assume that inferences can reflect, at different times, the operation of one set of mental processes (System1) or the other (System2). The fast, association-driven System1 is triggered whenever it encounters information it can process and is rather undemanding of cognitive resources. The analytic and reason-oriented System2 must be deliberately engaged and controlled, is slow, and demanding of capacity. System1 operates on contextualized tasks, taking into account semantic content and conversational principles. The operation of System2, in contrast, depends on the decontextualization of the task, and on the activation of abstract rules of inference. Dual-process accounts of conditional reasoning (Best, 2005; Klaczynski & Daniel, 2005; Klaczynski, Schuneman, & Daniel, 2004; Schroyens, Schaeken, & Handley, 2003) usually adopt a coarse-grained approach with respect to System1 and System2 processes. While students of conditional reasoning pay heed to the fact

Jean-François Bonnefon, Laboratoire Cognition, Langues, Langage et Ergonomie (CNRS, UTM, EPHE), Toulouse, France. Michael Eid, Free University of Berlin, Germany. Stéphane Vautier, Centre d'étude et de recherche en psychopathologie (UTM), Toulouse, France. Saïd Jmel, UMS838, Université de Toulouse le Mirail, France. This research was supported by a grant from the University of Toulouse. We thank Caroline Fourio for her technical assistance, David Over and one anonymous reviewer for helpful comments, and Lisa Trierweiler for her careful reading of the manuscript.

Correspondence concerning this article should be sent to Jean-François Bonnefon, Laboratoire Cognition, Langues, Langage et Ergonomie, Maison de la recherche, 5 allées A. Machado, 31058 Toulouse Cedex 9, France. E-mail: bonnefon@univ-tlse2.fr.

[1] Please note that every time we write of a 'correct' conclusion or answer, we mean an answer that is correct from the perspective of deductive logic. Similarly, when we speak of reasoners of high or low ability, we never mean more than high or low ability at giving logically correct answers. We do *not* mean anything as general as, e.g., intelligence.

that System1 derives conclusions from semantic content *or* pragmatic implicatures, the distinction between these two influences is seldom pushed further. Similarly, while it is acknowledged that System2 operates first by inhibiting System1 processes, *then* by recruiting abstract rules of inference, the consequences of this distinction are rarely considered (see, however, Evans & Over, 2004, chapter 9).

We suggest taking a more fine-grained approach to System1 and System2 processes in conditional reasoning. More precisely, we will argue that responses to conditional syllogisms are not solely determined by whether they are the result of System1 or System2 processing: Implicature-based, pragmatic System1 does not always yield the same answer as content-based, semantic System1. Likewise, the answer yielded by System2 will depend on whether it is only the result of the inhibition of System1 output, or whether an abstract rule of inference was actively recruited to generate a conclusion. This account has precise consequences for the measurement of conditional reasoning abilities. Not only it supposes continuous differences between individuals (as usual in the assessment of most cognitive achievements), but it also supposes structural, qualitative differences. Reasoners cannot simply be ordered on an ability continuum, but they have to be qualitatively compared with respect to their response process, that is, with respect to the reasoning subsystem that underlies their answers. Quantitative differences will then be found within each qualitative subpopulation of reasoners.

We will now give a detailed account of the various responses we expect from pragmatic System1, semantic System1, inhibitory System2, and generative System2 (see Table 1 for a summary). Then, we will seek evidence in the extant literature for the existence of these four subgroups. Next, we will present the analysis of the responses to 18 conditional syllogisms with the mixed Rasch model that exactly reflects our theoretical proposal.

Pragmatic or Semantic System1

Conversational pragmatics has been shown to affect conditional reasoning on a variety of problems (Bonnefon & Hilton, 2002, 2004; Bonnefon & Villejoubert, 2007; Stevenson & Over, 2001). In particular, pragmatics can affect even the simplest conditional reasoning problems by means of *invited inferences:* The pragmatic principles that govern conversation are such that the assertion of a conditional 'if *p*, then *q*' invites its converse 'if *q* then *p*' and its obverse 'if *not*−*p* then *not*−*q*' (Geis & Zwicky, 1971). These conversational implicatures do invite the determinate answer to DA and AC, but do *not* directly invite the MT inference from *not*−*q* to *not*−*p*. Pragmatic implicatures thus licence the determinate response to AC and DA, but not to MT. Hence, reasoners who rely solely on *pragmatic* System1 processing should, as a group, show high endorsement rates of the determinate answer to AC and DA, and a low endorsement rate of the determinate answer to MT. This group should be characterized by the 'all-wrong' individual pattern of responses: For a given triple of AC, DA, and MT syllogisms featuring the same conditional statement, reasoners applying pragmatic principles should most often give normatively incorrect answers to all three AC, DA, and MT.

Predictions are different with respect to reasoners who rely on the *semantic* System1, that is, individuals whose reasoning is affected by their background knowledge about the semantic content of the conditional. It has been repeatedly shown (Bonnefon & Hilton, 2002; Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991; Markovits & Quinn, 2002; De Neys, Schaeken, & d'Ydewalle, 2003; Politzer & Bonnefon, 2006; Thompson, 1994, 1995) that responses to AC, DA, and MT are affected by the retrieval of some specific type of information in semantic memory, and this influence is usually attributed to System1 processing (Evans, 2002; Evans & Over, 1996, 2004; Stanovich & West, 2000).[2] More precisely, endorsement of the determinate answer to DA and AC is negatively affected by the retrieval of alternatives (i.e., possible ways for the consequent of the conditional to occur when its antecedent is false), and endorsement of the determinate answer to MT is negatively affected by the retrieval of disablers (i.e., possible ways for the consequent to be false although the antecedent is true).

Semantic System1 reasoners will thus endorse determinate answers on AC, DA, and MT as a function of the number of alternatives and disablers that come to mind, depending on the conditional. Endorsement of AC and DA will not be as frequent as it is for the pragmatic System1 group—endorsement of MT, on the other hand, will be more frequent. No individual response pattern should be dominant in this group, as the response pattern will, for each conditional, be a function of the number of alternatives and disablers it evokes. Because the number of determinate responses depends on background knowledge, and because large individual differences in background knowledge can be expected, large quantitative individual differences in this subgroup are likely to be found.

Inhibitory or Generative System2

System2 is believed to operate on decontextualized premises, through the inhibition of background knowledge about their pragmatic and semantic aspects (Handley, Capon, Beveridge, Dennis, & Evans, 2005; De Neys et al., 2005; Simoneau & Markovits, 2003—see also Goel, Buchel, Frith, & Dolan, 2000; Goel & Dolan, 2003, for neuroimagery evidence). Inhibition of background knowledge will particularly discourage the endorsement of the determinate answer to AC and DA. The endorsement of the determinate answer to MT, on the other hand, will depend on whether System2 reasoners can recruit an abstract strategy for reductio ad absurdum (Best, 2005; Evans & Over, 2004); and not all reasoners will be sophisticated enough to do so.

Reasoners who only rely on the inhibition of background knowledge will thus (a) block the determinate answers to AC

[2] This might become a point of contention, as De Neys, Schaeken, and d'Ydewalle (2005) and Verschueren, Schaeken, and d'Ydewalle (in press) have recently considered the possibility that this influence might operate partly through conscious, cognitively demanding System2 processes.

Table 1
Characterization of the Four Groups by System of Reasoning, and Mechanism Engaged Within this System ('mod.' is the abbreviation of moderate).

	System1		System2	
	Pragmatic	Semantic	Inhibitory	Generative
Modal pattern	*all-wrong*	*none*	*all-blocked*	*all-correct*
Rate of correct AC	low	mod.	high	high
Rate of correct DA	low	mod.	high	high
Rate of correct MT	low	mod.	low	high
Ability	*low*	*low/mod.*	*mod./high*	*high*

and DA, and (b) endorse the undeterminate answer to MT, for lack of a strategy that would allow them to derive the determinate answer. As a consequence, the group of inhibitory System2 reasoners should be characterized by low endorsement rates of the determinate answer to AC, DA, and MT, and by the 'all-blocked' individual pattern of responses: For a given triple of AC, DA, and MT syllogisms featuring the same conditional statement, reasoners in this group should most often endorse the undeterminate answer to all three AC, DA, and MT. In psychometric terms, this means that these reasoners should find it more difficult to give a logically correct answer to MT syllogisms than to AC and DA syllogisms featuring the same conditional. From a quantitative perspective, mid to high ability scores are expected from this subpopulation.

Finally, the most sophisticated reasoners (who have access to an abstract reductio strategy) will reject the determinate answer to AC and DA (for they inhibit System1 responses), and endorse the determinate answer to MT (which was actively generated by reductio). Hence, generative System2 reasoners should be characterized by low endorsement rates of the determinate answer to AC and DA, and high endorsement rate of the determinate answer to MT. This group should also be characterized by the 'all-correct' individual response pattern: For a given triple of AC, DA, and MT syllogisms featuring the same conditional, reasoners in this group should most often endorse the normatively correct answer to all three, AC, DA, and MT. Hence, in psychometric terms, we do not expect differences between the difficulties of AC, DA, and MT syllogisms featuring the same conditional. Quite strikingly, that means that this group is *structurally equivalent* to the pragmatic System1 group, as no qualitative differences exist between the two groups. However, individuals in the generative System2 group should show high ability values, compared to the low ability values expected from the pragmatic System1 group. Thus, even though qualitative differences are not expected between the pragmatic System1 and generative System2 groups, these two groups should be clearly different in quantitative terms.

In summary, we expect quantitative and qualitative differences between individuals solving conditional reasoning tasks. Ideally, we expect three qualitative groups of individuals. In the first group, reasoners should find it equally difficult to give a logically correct answer to AC, DA, and MT. However, this group should split in two quantitatively distinct subgroups: (a) pragmatic System1 reasoners who systematically give the logically incorrect answer to all syllogisms, and show comparatively low abilities; and (b) generative System2 reasoners who systematically give the logically correct answer to all syllogisms, and show comparatively high abilities.

In the second group (semantic System1 reasoners), we expect responses to be largely influenced by the semantic content of the conditionals: Differences should be found in the difficulties of AC, DA, and MT, albeit not in a systematic way. Semantic System1 reasoners should have mid to low ability scores. Finally, in the third group (inhibitory System2 reasoners), we expect that MT will be solved less easily than AC and DA. Inhibitory System2 reasoners should have mid to high ability scores.

Prior empirical support

Our account is consistent with existing results that show a monotonic decrease of determinate responses to AC and DA in developmental studies (Barouillet, Grosset, & Lecas, 2000; Markovits, Fleury, Quinn, & Venet, 1998) or as a function of increasing cognitive capacity (Newstead, Handley, Harley, Wright, & Farelly, 2004). On the other hand, our account does not expect any simple monotonic relation to hold between cognitive sophistication and the rate of determinate responses to MT. Indeed, Newstead et al. (2004) observed a very weak linear correlation between endorsement of MT and cognitive ability, and Evans, Handley, Neilens, and Over (2006) observed that endorsement of MT appeared to sharply *decrease* with reasoning sophistication. Furthermore, a number of authors (Barouillet et al., 2000; Rumain, Connell, & Braine, 1983; O'Brien & Overton, 1980, 1982) have observed a puzzling developmental trend in the endorsement of MT: In broad agreement with our account, determinate responses to MT appear to increase and peak in preadolescent years, then to decrease with adult age. This developmental trend would correspond in our framework to a shift from pragmatic to semantic System1, then to inhibitory System2 in adult age. It is noteworthy that the studies we have mentioned so far do not report the second peak of MT endorsement we would expect with the shift from inhibitory System2 processing to generative System2 processing. We suspect that this might be due to sample size limitations, as the group of generative System2 reasoners might be too small to compensate for the larger group of inhibitory System2 reasoners.

In any case, these results only offer indirect support to our account. To make a direct case for our proposal, we need to move from correlational studies to classification studies. That is, instead of looking at bivariate correlations between responses to conditional syllogisms and measures of cognitive sophistication, we need a typological approach for classifying individuals with respect to the qualitative and quantitative differences they manifest when solving conditional syllogisms.

Contemporary psychometrics can provide us with the tool we need to conduct a qualitative classification, namely, latent class analysis (Goodman, 1974; Lazarsfeld & Henry, 1968; Rost & Langeheine, 1997). Latent class analysis can be used to identify distinct subpopulations from multivariate categorical data (in this case, answers to conditional syllogisms) and to assign individuals to their most likely subpopulation.

To the best of our knowledge, only two studies have applied latent class analysis to conditional reasoning. Rijmen and De Boeck (2003) analysed responses of high-school students to complex conditional reasoning problems, combining one of the four elementary inferences (AC, DA, MP, MT) with another component such as conjunction or disjunction. A first analysis was run on AC- and DA-based problems, which yielded two latent classes; these classes were interpreted as corresponding to a material vs. biconditional interpretation of conditional statements. A separate analysis was conducted on MP- and MT-based problems. This analysis again yielded two classes. Members of the second class had relatively more difficulties with MT, but also higher general propositional reasoning ability. Spiel, Gittler, Sirsch, and Glück (1997) gave early adolescents a set of eight conditional syllogisms to solve, corresponding to one abstract instance and one concrete instance of AC, DA, MP, and MT. Latent class analysis yielded two classes that mainly differ in the rate of determinate responses to AC and DA.

These two studies revealed important insights into the cognitive stages early and late adolescents go through with respect to conditional reasoning. Both studies show that there are strong qualitative differences between individuals. However, they assume that all quantitative differences can be explained by the two classes, since ability differences are not allowed within classes. In other words, they both assume that all interindividual differences can be explained by a dichotomy in ability.

Additionally, using two conditional statements only, as in the study of Spiel et al. (1997), seems insufficient to control for content effects. Finally, since our predictions concern the general pattern of answers to AC, DA, and MT syllogisms, we need to run one general latent class analysis, rather than the two separate analyses of Rijmen and De Boeck (2003). The study we now report was designed and conducted to deal with these concerns. Our study goes beyond previous important empirical works by (a) integrating qualitative and quantitative individual differences, (b) developing and testing a general model comprising all three syllogisms, and (c) allowing the analysis of content effects. Instead of applying latent class analysis, we will use mixed Rasch models, an integration of latent class analysis and the Rasch one-parameter logistic test model, that allow the consideration of qualitative and quantitative differences simultaneously.

Methods

Participants, material, and procedure

Participants were recruited by third-year psychology students as a course requirement. Each student made a list of several men and women who were older than 18, not studying psychology, and willing to take part in a survey on reasoning—no other restriction applied, e.g., family members were permitted. Each student then randomly selected one male and one female participant from this list. It was expected that this recruitment procedure would promote variety in age, occupation, and education, while ensuring equal proportions of male and female participants. No incentive was offered to participants. In the rare cases when a randomly selected participant did not consent to take part in the survey, the student made a second random selection from his or her list.

Of the 486 participants who returned a fully completed questionnaire (49% men, 51% women, mean age = 31, *SD* = 12.6), 20% had completed graduate school or an equivalent school form, 41% had the equivalent of an undergraduate education, 25% graduated from high school only, and the educational level of 14% was lower than high school. The sample included a large proportion of students (37%), but the remaining 63% came from practically all professional perspectives (including 10% unemployed).

The conditional reasoning task consisted of six blocks of three syllogisms. All blocks comprised one AC, one DA, and one MT syllogism (presented in a different order in each block). These syllogisms were embedded in a simple context that was different for each block. Here is one example of a complete block:

> You are a doctor in a tropical country. According to your experience, *if a patient has malaria, he makes a quick recovery*.
>
> **Modus Tollens** You observe the following situation: A patient does not make a quick recovery. Does the patient have malaria? ('Yes', 'No', 'Maybe')
>
> **Denying the Antecedent** You observe the following situation: A patient does not have malaria. Does the patient make a quick recovery? ('Yes', 'No', 'Maybe')
>
> **Affirming the Consequent** You observe the following situation: A patient makes a quick recovery. Does the patient have malaria? ('Yes', 'No', 'Maybe')

The other five conditionals (all taken from Thompson, 2000) were: 'If there is a low pressure system, it will rain;' 'If a restaurant sells liquor, it must have a liquor license;' 'If someone has broken an item in the store, they must pay for it;' 'If a company makes a big profit, the price of their shares will go up;' 'If the content of the bottle is poisonous, it must be labelled “poison”.’ Participants filled out the questionnaire at their own pace and at their own place, under the supervision of the student who recruited them. The survey was conducted in French.

Methods of data analysis

Each response to each syllogism was coded ‘1’ when the response was logically correct and ‘0’ when it was logically incorrect. The responses of all individuals to all 18 questions were analysed with the mixed Rasch model (Rost, 1990; von Davier & Carstensen, in press) using the computer program WINMIRA (von Davier, 2001).

The mixed Rasch model is an extension of the Rasch model (Rasch, 1960/1980) and the latent class model. As a Rasch model, it assumes that the probability of a correct response depends on the ability of an individual and the difficulty of an item. The ability of an individual is represented by a person parameter θ_v that indicates the standing of an individual on a latent continuous ability variable. The higher the person parameter of an individual, the higher is the ability of that individual. The difficulty of an item is represented by a difficulty parameter σ_i that indicates the standing of the item on the latent ability variable. The value of the difficulty parameter corresponds to a value on the latent ability variable for which the probability to solve the item is .50. The probability of a correct response is a non linear function of the latent ability variable and depends on the difference between the ability of an individual and the difficulty of an item. If this difference is positive, the probability to solve the item is higher than .50; if the difference is 0, the probability equals .50; and if the difference is negative, the probability is smaller than .50. Formally, the Rasch model is defined by the following equation:

$$P(X_{vi} = 1) = \frac{e^{(\theta_v - \sigma_i)}}{1 + e^{(\theta_v - \sigma_i)}} ,$$

where X_{vi} denotes the response variable of an individual v on an item i, which takes the value 1 in case of a correct answer and the value 0 in case of an incorrect answer.

The item characteristic curve has the same form for all items and depends only on the difficulty parameter. The Rasch model assumes that there is only one latent ability variable that can explain the item responses (assumption of unidimensionality). Furthermore, it assumes that all associations between the observed responses are explained by the latent variable. When correcting for ability differences, there should be no further associations between the items (assumption of local independence). The aim of a Rasch analysis is to test whether these assumptions are true and to estimate the person parameters and the item difficulties. The Rasch model is based on strong assumptions that might often be violated. One strong assumption is that the item difficulties do not differ between individuals and are the same for all members of the population. That means that the ordering of the items with respect to their difficulties do not differ between individuals. This assumption does not hold if there are qualitative differences between individuals, that is, when the relations of the item difficulties differ between individuals. One item can be easier than another item for one individual but more difficult for another individual. This is particularly the case if individuals use different solution strategies, and if the solutions of the items are prone to these differences in solution strategies. In order to overcome this problem, the Rasch model has been extended to the mixed Rasch model, which is a combination of the Rasch model with the latent class model.

As a latent class model, the mixed Rasch model assumes that the population is not homogeneous but consists of G different nonoverlapping subpopulations that differ with respect to the response probabilities of the items. Whereas the latent class model assumes that there are no individual differences within a class with respect to the response probabilities, the mixed Rasch model allows individual differences within classes. In fact, the mixed Rasch model assumes that within each latent class, a Rasch model holds, but that the values of the parameters of the model can differ between classes. This means that in each class, the probability of the correct response depends on the ability of an individual and the difficulty of an item, but that, for example, the item difficulties can differ between classes. Formally, the mixed Rasch model is defined by the following equation (Rost, 1990):

$$P(X_{vi} = 1|g) = \frac{e^{(\theta_{vg} - \sigma_{ig})}}{1 + e^{(\theta_{vg} - \sigma_{ig})}} ,$$

where X_{vi} denotes the response variable of an individual v on an item i, which takes the value 1 in case of a correct answer and the value 0 in case of an incorrect answer.

According to this model, the probability P of a correct response to an item i from an individual v belonging to a latent class g depends on the difference between the ability θ_{vg} of this individual and the difficulty parameter σ_{ig} of that item in class g. Hence, the item parameters can differ between classes, representing structural differences in the response process. The aim of a mixed Rasch analysis is to estimate the person parameters and the item difficulties as well as the class sizes π_g and the latent person parameter probabilities.

The latent classes are disjoint and exhaustive. That means that each individual must belong to one latent class and can only belong to one latent class. Because of the disjoint, nonoverlapping classes, the probabilities π_gs of the latent classes (i.e., class sizes) must add up to one. It is not known beforehand to which latent class an individual belongs—however, for each individual, the probability of belonging to each of the latent classes can be estimated based on the response pattern of this individual and the parameters of the model. These probabilities add up to one for each individual. Moreover, for each individual an ability parameter is estimated for each of the different classes. An individual is assigned to the latent class for which her assignment probability is maximal. The mean of all assignment probabilities of individuals assigned to a class can be considered as the assignment reliability.

After having assigned individuals to the class with the highest assignment probability, the class membership can be considered as a nominal scaled variable, and can be related to other variables to analyse the way in which the classes differ with respect to other external variables. It is important to note that the number of latent classes is not a parameter of the model that can be estimated. In contrast, the fits of several mixed Rasch models with different numbers of latent classes must be compared to find the optimal number of classes. The final number of latent classes is given by the number of latent classes of the best-fitting mixed Rasch model.

There are several indicators for evaluating the fit of a mixed Rasch model. As for many other models of categorical data the observed frequencies of the observed response vectors can be compared with the expected frequencies of these vectors given by the mixed Rasch model. There are several test statistics for comparing the expected and observed frequencies (Read & Cressie, 1988), among which the χ^2 distribution and the likelihood ratio test are the most widely used. These test statistics are distributed according to a χ^2 distribution when the expected frequencies of each possible response pattern are at least 1 (Rost, 1990). Given that we will be analyzing 18 binary items, there are 262,144 possible response patterns. Hence, we cannot trust that these test statistic are really distributed according to a χ^2 test in the current application. In this case, an estimation of the p-value can be calculated with the bootstrap method. However, as von Davier (1997) has shown, the bootstrap methods only works fine for bootstrapping the distribution of the Pearson χ^2 test and the Cressie-Read test, but not for others like the likelihood ratio test. Therefore, we will only consider these two test statistics for evaluating our model.

These fit coefficients can only be applied to test the fit of a model with a given number of classes. They can not be used to compare several mixed Rasch models that differ in the number of classes. Furthermore, the likelihood ratio test can not be applied to compare the fit of two mixed Rasch models differing in the number of classes because one important regularity condition is violated. If, for example, the fit of a 3-class model is compared with the fit of a 2-class model, the 2-class model is a special case of the 3-class model with one class probability equal to 0. But if one parameter is fixed to a boundary value of the parameter space, the likelihood ratio test cannot be applied. However, in order to compare the fit of different models, information criteria like Akaike's information criterion and the Bayesian information criterion can be used. These criteria compare the general fit of a model with the number of parameters estimated. Different models can be compared according to their values on the information criteria. The best model is the model with the lowest values on the information criteria; this is the model that shows a good fit with the lowest number of parameters. According to Rost (2004), the Bayesian information criterion is preferable to Akaike's information criterion in the case of a large number of response patterns, and it will thus be used here. From a theoretical point of view, we expect a 3-class model. In order to evaluate the relative fit of this model, we will compare it with a 1-class model (Rasch model), a 2-class model, and a 4-class model. We will use the Bayesian information criterion coefficient to examine whether the 3-class model has a better fit than the models with more or less classes. If this can be confirmed, we will test the fit of the model with respect to the Pearson χ^2-test and the Cressie-Read test using bootstrap analysis.

Results and discussion

The Bayesian information criterion (BIC) values for the Rasch model and the mixed Rasch models with 2, 3, and 4 classes are 8506, 8467, 8465, and 8492, respectively. According to the BIC, the model with three classes is the best fitting model. However, its BIC value does not differ much from the model with two latent classes. Hence, we will compare the 3-class model with the 2-class model to learn more about the differences between the two solutions, and whether a third class is necessary. The model with four latent classes shows a higher BIC value that indicates that a fourth class is not necessary. Since the fourth class is rather small (9%), and because we want to avoid spurious classes, we will concentrate on the 2- and 3-class solutions.

The 2-class model consists of a larger class (67%) and a smaller class (33%). The class sizes of the 3-class solution are 45%, 35%, and 20%. A careful comparison of the estimated difficulty parameters reveals that the profile of the item parameters of the largest class of the 2-class solution is very similar to the difficulty parameters in the second largest class of the 3-class solution. Moreover, the profiles of the item parameters in the smallest classes of the 2- and 3-class solutions are very similar. The largest class of the 3-class solution has no counterpart in the 2-class model. Because the 3-class model covers the structural differences that are found in the 2-class model, and because the third class, that is not present in the 2-class model, is the largest class, we decided to present the 3-class model.

This decision is also supported by the bootstrap analysis. The 3-class solution shows a better fit with respect to the p-values of the bootstrap analyses: The p-values of the Pearson χ^2 test and the Cressie-Read test for the 3-class model are 0.09 and 0.03, respectively, showing that this model fits the data rather well. The p-values of the Pearson χ^2 test and the Cressie-Read test for the 2-class model are 0.01 and 0.003, respectively, showing that the fit of the 2-class model is worse than the fit of the 3-class model. These p-values show that the 3-class model has to be rejected according to an α of 0.05 and with respect to the Cressie-Read statistic but not the Pearson test. The 3-class model does not have to be rejected with respect to an α of 0.01. However, we will not choose a strict α level for deciding for or against a model. It is well known from latent variable modelling that good models will be rejected when the sample size is large because of a large power. In line with other latent variable approaches such as structural equation modelling we consider the p-value as a measure of the goodness of fit of a model that has to be considered as one of several fit coefficients. The bootstrap p-values indicate a good fit of the 3-class model and a less good fit of the 2-class model.

Table 2 displays the rates of correct responses to AC, DA, and MT in the three classes. The graphical output of the model is shown in Figure 1. The left panel of Figure 1 shows the

Figure 1. Structural and quantitative differences between the three latent classes. The left panel shows how difficult each item is for individuals the in three latent classes: The higher the parameter, the greater the difficulty. Items are arranged by blocks of AC, DA, and MT arguments that feature the same content. For simplicity reasons the AC, DA, and MT tasks are only indicated for the first block. For all other blocks only the MT task is indicated, the sequence AC–DA–MT is always the same. The right panel shows the estimated raw score distributions (and the corresponding estimates of the latent ability parameter, from −8 to +7) in the three classes.

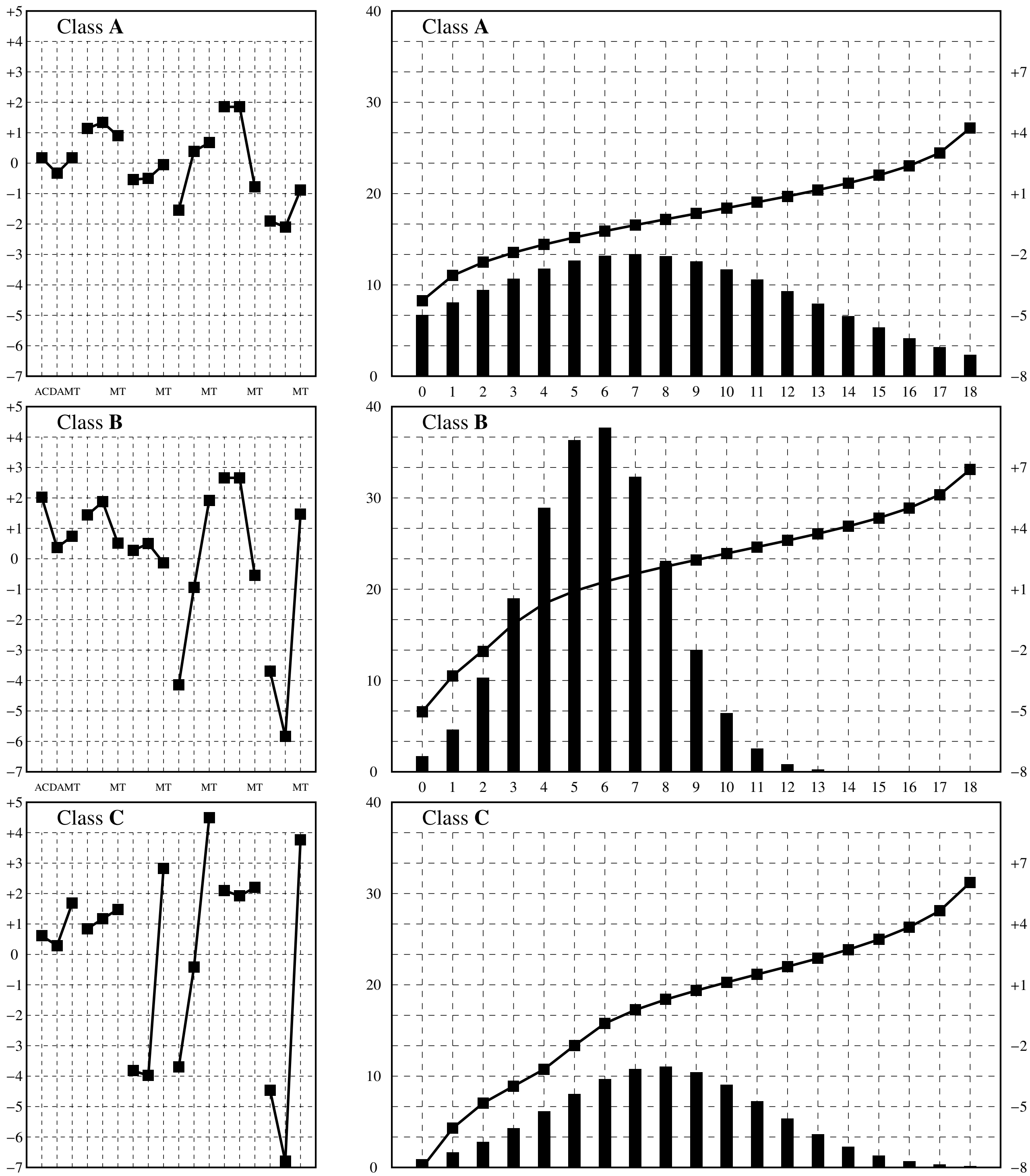

Table 2
Correct Response Rates (in % of Answers) to the Conditional Arguments in the Three Latent Classes.

	CLASS A	CLASS B	CLASS C	*Overall*
AC	38	47	63	*46*
DA	33	43	60	*41*
MT	20	45	15	*28*

difficulty parameters of each item in each class: The difficulty parameter of an item indicates the value on the latent variable where the probability of a correct response equals .50. The higher the value of a difficulty parameter the higher is the ability needed to give a correct answer. Strong differences exist between the three groups, mainly in how sensitive they are to argument type. The right panel of Figure 1 shows the distribution of estimated raw scores in each latent class, together with the corresponding estimation of latent score: The (estimated) raw score of an individual corresponds to the (estimated) number of logically correct answers given out of 18 answers total. The corresponding latent score is the raw score corrected for measurement error. The latent score is a monotonically increasing yet nonlinear function of the raw score.

Class A (35% of all individuals) shows comparatively small differences in the difficulty parameters of all syllogisms. In particular, all arguments in a block featuring the same conditional have roughly the same difficulty. Only for the shop context (If someone has broken an item in the store, they must pay for it) is AC a bit easier than DA and MT. For the poison context (If the content of the bottle is poisonous, it must be labelled 'poison'), MT is somewhat easier than the two other syllogisms. The estimated raw score distribution in this class covers the whole range of abilities. This class is closely in line with the first group we expected on theoretical grounds. Qualitatively speaking, reasoners in this class are largely uninfluenced by argument or content. Quantitatively speaking, this class comprises both individuals with very low and very high raw scores (i.e., individuals who get most answers wrong and individuals who get most answers right, respectively), which is not the case for the other classes. All these results strongly suggest that this class is, as expected, a mix of pragmatic System1 reasoners and generative System2 reasoners.

Class B (45% of all individuals) is the class for which the content of conditionals has the largest influence: The difficulty parameters of the syllogisms show the greatest variations from one block to another. MT is solved less easily than DA or AC in the shop and malaria contexts, but the opposite is true in the poison context. Furthermore, DA and AC are solved especially easily in the malaria context. In quantitative terms, raw scores in this class are symmetrically distributed and in the lower half of the distribution. Overall, strong influence of content rather than argument type plus low ability scores suggest that this class corresponds to the hypothesized group of semantic System1 reasoners.

Class C (20% of all individuals) is the class for which argument type has the strongest influence: Huge differences are found in the difficulty of syllogisms featuring the same content as a function of whether they are of the AC, DA, or MT type. A careful examination of these differences reveals that there are comparatively small differences between AC and DA syllogisms, but strong differences between MT and the other two syllogisms. In all cases, MT syllogisms are more difficult than the two other syllogisms. This class is closely in line with the proposed subpopulation of inhibitory System2 reasoners. In qualitative terms, individuals in this class find it easy to block the incorrect answer to AC and DA, and yet extremely difficult to produce the correct answer to MT. In quantitative terms, raw scores in this class are symmetrically distributed and higher than raw scores observed in Class B, which we have taken to correspond to semantic System1 reasoners.

General Discussion

Proponents of dual-process accounts of conditional reasoning have warned against an overly simplifying assumption (Evans & Over, 2004; Klaczynski, 2001): One should not think of logically valid answers to be solely the result of System2 processes, and respectively, logically invalid answers to be solely the result of System1 processes. In this article, we have pointed to another overly simplifying assumption: Different mechanisms operate both within System1 and System2, and these different mechanisms can very well yield different answers to a given conditional syllogism. The case of MT is a striking illustration of this point. The determinate answer to MT is not produced by pragmatic System1, but it is (sometimes) produced by semantic System1. Only then it is no longer produced by inhibitory System2—but it comes back full-force with generative System2.

Untangling the answers yielded by pragmatic and semantic System1 (on the one hand), and inhibitory and generative System2 (on the other hand), we have hypothesized the existence of four subgroups and three latent classes of which one latent class comprises two subgroups. These four subgroups are differentiated both from a qualitative and a quantitative perspective resulting in a mixed Rasch model with three classes and quantitative differences within classes.

These results clearly show that the content and context effects which form the bulk of System1 processing need not and ought not be considered as inextricably interwoven. Pragmatic and semantic effects on conditional reasoning are separable. Some reasoners simply generate conclusions from pragmatic implicatures, while others base their answers on prior knowledge about the world; and these different response processes are revealed in a latent class analysis. Similarly, results suggest that reasoners can inhibit System1 responses without having actively generated a System2 answer. In particular, some individuals, who cannot recruit the abstract reductio strategy for MT, will nevertheless inhibit content and context effects. This suggests that the *conflict* between System1 and System2 outputs (Sloman, 1996) is not always a prerequisite for System2 to override System1.

Reasoners may attempt to block a System1 response even though they have failed to generate an abstract, rule-based System2 response.

Mental Models

We have argued that a fine-grained approach to System1 and System2 mechanisms is needed to advance our understanding of conditional reasoning. Yet, might our results be explained without appealing to any dual-process approach at all? That is, can the different subgroups of reasoners we found be explained within a general, single-process account of conditional reasoning such as mental model theory?[3]

From the perspective of the mental model theory of conditional reasoning (Johnson-Laird & Byrne, 2002; see also Bonnefon, 2004; Evans, Over, & Handley, 2005), conclusions of conditional syllogisms are read from the set of mental models that was generated during premise interpretation. Each of our subgroups of reasoners is characterized by its unique pattern of answers to conditional syllogisms; as a consequence, a mental model account of our results would assume that each subpopulation generated a distinct set of mental models. More precisely, individuals we call inhibitory System2 reasoners would assign conditional statements a 'tautological' interpretation (models pq, $\neg pq$, $p\neg q$, and $\neg p\neg q$, where $\neg p$ represents the negation of p), while individuals we call generative System2 reasoners would assign conditional statements a 'conditional' interpretation (models pq, $\neg pq$, and $\neg p\neg q$). Semantic System1 reasoners would not systematically assign conditional statements a given interpretation, but would apply 'pragmatic modulation' to incorporate prior knowledge into their mental models. Finally, pragmatic System1 reasoners would (paradoxically) appear to eschew pragmatic modulation entirely, and systematically represent conditional statements with the set of models pq, $p\neg q$, and $\neg p\neg q$.

The main problem with this account relates to this last subgroup. The sets of models pq, $p\neg q$, and $\neg p\neg q$ has already been identified by Johnson-Laird and Byrne (2002) as corresponding to an 'enabling' interpretation of the conditional. The enabling interpretation applies when the antecedent p is necessary for the consequent q to occur, that is, when p is the only enabling condition for q, or, in other terms, when q cannot occur if p is not satisfied. Johnson-Laird and Byrne (2002) give several examples of such conditionals: 'If you log on to the computer, then you may be able to receive e-mail;' 'If oxygen is present, then there may be a fire;' 'If it's her book, then she is allowed to give it away.' It seems highly implausible, however, that a large subpopulation of reasoners will almost systematically interpret any given conditional as expressing such an enabling condition.

Conditional Probabilities

An integral component of the dual-process theory of conditional reasoning (Evans & Over, 2004) is to consider conditionals in terms of *conditional probabilities.* According to this view, the degree of belief in the conclusion of a conditional syllogism critically depends on the degree of belief in the conditional statement 'if p then q'—and this degree of belief is adequately represented by the conditional probability $P(q|p)$ (Evans, Handley, & Over, 2003; Oberauer & Wilhelm, 2003; Over, Hadjichristidis, Evans, Handley, & Sloman, in press). In parallel, other authors suggested that the willingness to accept the conclusion of a conditional syllogism depends on the conditional probability of the conclusion given the minor premise (Liu, 2003; Liu, Lo, & Wu, 1996; Oaksford, Chater, & Larkin, 2000): E.g., the willingness to endorse DA would depend on $P(\neg q|\neg p)$.

Although we have not addressed this aspect of dual-process theories in this article, it can be readily integrated to our proposal. Consider for example the case of semantic System1 reasoners, who are essentially influenced by background knowledge related to the contents of the conditional. This background knowledge impacts their degree of belief in the conditional and the relevant conditional probabilities, which in turn impact their willingness to endorse the conclusion of conditional syllogisms (Weidenfeld, Oberauer, & Hörnig, 2005).

Only the semantic System1 subgroup is expected to be largely influenced by the contents of the conditional. Does that mean that the conditional probability $P(q|p)$ will not play a role in the reasoning of other subgroups? Not necessarily. Indeed, the probability $P(q|p)$ depends on other factors than just the semantic contents of p and q. In particular, pragmatic aspects of the situation, such as *who* asserted the conditional, can impact $P(q|p)$ and the subsequent willingness to endorse the conclusion of conditional syllogisms. Consider one of the statement we used in our study: 'If a patient has malaria, he makes a quick recovery.' As shown in Stevenson and Over (2001), the conditional probability attached to this statement is higher when it is asserted by an expert doctor than when it is asserted by a medical student. Pragmatic System1 reasoners, even though they might be insensitive to the exact semantic contents of the conditional, are likely to be influenced by such contextual factors.

Conclusion

We hope to have made a convincing case for our approach, which mixes fine-grained dual-process theorization with a mixed Rash model methodology. We are, nevertheless, aware that we have so far only addressed one aspect of conditional reasoning, that is, conditional syllogisms. We plan to extend this approach to the selection task and the truth-table task—with mixed expectations. We believe, along with Thompson (2000), that most of the interpretative processes that influence conditional reasoning are task specific; and thus, that subpopulations of reasoners on a conditional reasoning task will not readily map onto, for example, subpopulations of reasoners on the selection task. On the other hand, Newstead et al. (2004) report some association between general intelligence, the tendency to resist AC and DA,

[3] Simply because we qualify mental model theory as a single-process account does not preclude the possibility to include it as the analytical component *within* a dual-process account (Verschueren et al., in press).

and the consistency with which one gives a normative answer to the indicative selection task. Moreover, Evans et al. (2006) report promising results relating cognitive sophistication, responses to conditional syllogisms, and responses to the truth-table task. Only time (in the form of future research) will tell whether these associations are really the tip of the psychometric iceberg.

References

Barouillet, P., Grosset, N., & Lecas, J. F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition*, *75*, 237–256.

Best, J. (2005). Recognition of proofs in conditional reasoning. *Thinking and Reasoning*, *11*, 326–348.

Bonnefon, J. F. (2004). Reinstatement, floating conclusions, and the credulity of mental model reasoning. *Cognitive Science*, *28*, 621–631.

Bonnefon, J. F., & Hilton, D. J. (2002). The suppression of modus ponens as a case of pragmatic preconditional reasoning. *Thinking and Reasoning*, *8*, 21–40.

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

Bonnefon, J. F., & Villejoubert, G. (2007). Modus Tollens, Modus Shmollens: Contrapositive reasoning and the pragmatics of negation. *Thinking and Reasoning*, *13*, 207–222.

Cummins, D. D. (1995). Naïve theories and causal deduction. *Memory and Cognition*, *23*, 646–658.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory and Cognition*, *19*, 274–282.

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

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking and Reasoning*, *11*, 349–381.

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, *128*, 978–996.

Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2006). *The mental representation of abstract conditional sentences: A study of qualitative individual differences.* Manuscript submitted for publication.

Evans, J. St. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 321–335.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

Evans, J. St. B. T., & Over, D. E. (2004). *If*. Oxford: Oxford University Press.

Evans, J. St. B. T., Over, D. E., & Handley, S. J. (2005). Suppositions, extensionality, and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review*, *112*, 1040–1052.

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

Goel, V., Buchel, C., Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage*, *12*, 504.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition*, *87*, B11–B22.

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

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. T. (2005). Working memory, inhibitory control, and the development of children's reasoning. *Thinking and Reasoning*, *10*, 175–195.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 646–678.

Klaczynski, P. A. (2001). Analytic and heuristic processing influences on adolescent reasoning and decision making. *Child Development*, *72*, 844–861.

Klaczynski, P. A., & Daniel, D. B. (2005). Individual differences in conditional reasoning: A dual-process account. *Thinking and Reasoning*, *11*, 305–325.

Klaczynski, P. A., Schuneman, M. J., & Daniel, D. B. (2004). Theories of conditional reasoning: A developmental examination of competing hypotheses. *Developmental Psychology*, *40*, 559–571.

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

Liu, I. M. (2003). Conditional reasoning and conditionalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 694–709.

Liu, I. M., Lo, K. C., & Wu, J. T. (1996). A probabilistic interpretation of "if-then". *Quarterly Journal of Experimental Psychology*, *44*, 828–844.

Markovits, H., Fleury, M. L., Quinn, S., & Venet, M. (1998). The development of conditional reasoning and the structure of semantic memory. *Child Development*, *69*, 742–755.

Markovits, H., & Quinn, S. (2002). Efficiency of retrieval correlates with 'logical' reasoning from causal conditional premises. *Memory and Cognition*, *30*, 696–706.

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology*, *57A*, 33–60.

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

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

O'Brien, D., & Overton, W. F. (1980). Conditional reasoning following contradictory evidence: A developmental analysis. *Journal of Experimental Child Psychology*, *30*, 44–60.

O'Brien, D., & Overton, W. F. (1982). Conditional reasoning and the competence-performance issue: A developmental analysis of a training task. *Journal of Experimental Child Psychology*, *34*, 274–290.

Over, D. E., Hadjichristidis, C., Evans, J. St. B. T., Handley, S. J., & Sloman, S. A. (in press). The probability of causal conditionals. *Cognitive Psychology*.

Politzer, G., & Bonnefon, J. F. (2006). Two varieties of conditionals and two kinds of defeaters help reveal two fundamental types of reasoning. *Mind and Language*, *21*, 484–503.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, MA: University of Chicago Press. (Original work published 1960)

Read, T. R. C., & Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data*. NY: Springer Verlag.

Rijmen, F., & De Boeck, P. (2003). A latent class model for individual differences in the interpretation of conditionals. *Psychological Research*, *67*, 219-231.

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

Rost, J. (2004). *Lehrbuch testtheorie, testkonstruktion* [Textbook test theory, test construction] (2 ed.). Bern, Switzerland: Huber.

Rost, J., & Langeheine, R. (1997). A guide through latent structure models for categorical data. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 13–37). New York: Waxmann.

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.

Schroyens, W., Schaeken, W., & Handley, S. (2003). In search of counter-examples: Deductive rationality in human reasoning. *Quarterly Journal of Experimental Psychology*, *56A*, 1129–1145.

Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: Evidence for the role of inhibition and retrieval. *Developmental Psychology*, *39*, 964–975.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Spiel, C., Gittler, G., Sirsch, U., & Glück, J. (1997). Application of the Rasch model for testing Piaget's theory of cognitive development. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 108–114). New York: Waxmann.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645–665.

Stevenson, R. J., & Over, D. E. (2001). Reasoning from uncertain premises: Effects of expertise and conversational context. *Thinking and Reasoning*, *7*, 367-390.

Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory and Cognition*, *22*, 742–758.

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

Thompson, V. A. (2000). The task-specific nature of domain-general reasoning. *Cognition*, *76*, 209–268.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (in press). A dual-process specification of causal conditional reasoning. *Thinking and Reasoning*.

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*. http://www.mpr-online.de.

von Davier, M. (2001). WINMIRA 2001 [Computer software]. Kiel: Institut für die Pädagogik der Naturwissenschaften.

von Davier, M., & Carstensen, C. (in press). *Multivariate and mixture distribution rasch models*. NY: Springer Verlag.

Weidenfeld, A., Oberauer, K., & Hörnig, R. (2005). Causal and noncausal conditionals: An integrated model of interpretation and reasoning. *Quarterly Journal of Experimental Psychology*, *58A*, 1479–1513.