Bayesian Logic and Trial-by-Trial Learning

Momme von Sydow (momme.von-sydow@psychologie.uni-heidelberg.de)
Klaus Fiedler (klaus.fiedler@psychologie.uni-heidelberg.de)
University of Heidelberg, Department of Psychology, Hauptstr. 47-51,
D-69117 Heidelberg, Germany

Abstract

Standard logic and probability theory are both beset with fundamental problems if used as adequacy criteria for relating logical propositions to learning data. We discuss the problems of exception, of sample size, and of inclusion. Bayesian pattern logic ('Bayesian logic' or BL for short) has been proposed as a possible rational resolution of these problems. BL can also be taken as psychological theory suggesting frequency-based conjunction fallacies (CFs) and a generalization of CFs to other logical inclusion fallacies. In this paper, this generalization is elaborated using trial-by-trial learning scenarios without memory load. In each trial participants have to provide a probability judgment. Apart from investigating logical probability judgments in this trial-by-trial context, it is explored whether under no memory load the propositional assessment of previous evidence has an influence on further probability judgments. The results generally support BL and cannot easily be explained by other theories of CFs.

Keywords: Conjunction fallacy, probability judgments, trial-by-trial learning, Bayesian induction, logical predication.

Standard Logic and Probability Theory as Criteria for True Logical Propositions

The relationship between general logical propositions (or sentences) and evidence is fundamental to both epistemology and psychology. We here investigate general predication of logical relationships between two dichotomous attributes (or predicates), like "ravens are black *and* they can fly" (with the conjunction 'and'). What would be an adequate justification for such a type of sentences?

Arising from an old tradition going back to Aristotle, modern formal logic uses truth table definitions for all 16 logical connectives. The truth table definition may be used as a deterministic criterion of truth for empirical relationships. With regard to a conjunctive predication, like "ravens (*R*) are black (*A*) and they can fly (*B*)" ($A \wedge B|\ R$), the whole sentence is true (or, more correctly, 'not false') as long as one has observed only exemplars corresponding to true cells of a truth table (for the conjunction this is the '*a*-cell', '*A* & *B*'). In contrast, the proposition would be falsified, if one observed a single case defined to be false (here: *b*-cell: '*A* & $\neg B$'; c-cell: '$\neg A$ & *B*', or d-cell: '$\neg A$ & $\neg B$').

Problem of Exceptions Exceptions may not prove the rule, but in ordinary language exceptions are indeed regularly tolerated. This may reflect the deeper epistemological point that in the empirical world deterministic relationships are rather the exception than the rule. Actually, in philosophy of science it has been argued that strict falsificationism would absurdly imply that *all* important theories would be falsified. Even more so in normal language, as evident from our deterministic example, there exist exceptions: white (albino) ravens as well as ravens that cannot fly. If exceptions are the rule for contingent, empirical relationships, it seems reasonable to replace the strict deterministic truth criteria of logic by a high-probability criterion (see Schurz, 2005): *P*(black & can fly | ravens) > Ψ, with Ψ > .5. However, the following two problems beset a simple extensional probability criterion of truth as well as one based on standard formal logics.

Problem of Sample Size If we had to access the truth of "ravens are black and they can fly" without previous knowledge about ravens, either one confirmatory raven (*A* & *B* case) or many cases both equally yielded the same extensional probability of 1 (the number of confirmative cases divided by all cases). In the latter case, however, a higher subjective probability of this sentence seems justified. Therefore, a kind of second order probability, a probability concerning probabilities, is needed, as introduced in the model.

Problem of Inclusion The extension (all cases falling into a set) of a subset can never be larger than that of a superset. Comparing conjunctions and inclusive disjunctions, it follows that *P*(ravens are black AND they can fly) ≤ *P*(ravens are black OR they can fly or both) [formally: $P(A \wedge B|R) \leq P(A \vee B|R)$]. If we use extensional probabilities as truth criterion, the second sentence can therefore never be 'less true' than the first one. If one assumes at least some exceptions, the latter is even 'truer' in principle. Going one step further, the logical tautology, allowing for all values ("Ravens are black or not, and they can fly or not"), is *a priori* the extensionally most probable sentence [$P(A \vee B|R) \leq P(A\ \mathrm{T}\ B|R)$ or even $P(A \vee B|R) < P(A\ \mathrm{T}\ B|R)$]. Using standard (extensional) probabilities as truth criterion, one would therefore always have to choose tautologies as the most suitable hypothesis, regardless of the evidence and of the properties in question. In conclusion, if a truth criterion should be informative about the observable world, simple extensional probabilities in principle cannot provide a reasonable truth criterion.

Bayesian Logic

Bayesian pattern logic (or 'Bayesian logic', BL, for short) formulates a second order probability that given data may have been generated by noisy-logical patterns of probabilities. The model provides a technical, rational solution to the three mentioned problems and – in approximation – a potential psychological model of human induction of noisy-logical relationships as well. The model is part of a renaissance of Bayesian approaches in cognitive science (e.g., Chater, Tenenbaum, Yuille, 2006; Oaksford & Chater, 2007, Kruschke, 2008). The following sketch is meant to clarify the main idea of Bayesian logic (for more detail, see von Sydow, 2011).

The construction of the model starts with all 16 dyadic logical connectives known from standard propositional logic. The logical truth tables are taken as explanations that are distinguished from the data level. While standard logic makes no assumptions about probabilities of true classes in a truth table, Step 1 of BL formulates ideal explanations by assuming equi-probability of all true classes of a truth table. For instance, for the exclusive disjunction (*X* are *either A or B, but not both*) it is assumed that *P*(b-cell) = *P*(c-cell) = ½ (for no noise, *R* = 0). Thereby, 2 by 2 truth tables become 2 by 2 probability tables. Note, however, that such ideal explanations need not generate ideal data patterns. In Step 2 (cf. Fig. 1) the idea of exceptions is modeled by introducing possible levels of noise. For each possible level a uniform noise function is added to all four cells of probability table, followed by a normalization, so that the resulting sum of all four cells of a probability table adds up to unity. This results in a field of ideal (explanatory) noisy-logical patterns of probabilities, each with an additional second order probability: *P*(*A* connective *B*, noise level *R* | data)) =: *P*(*A* o *B*, *R*| *D*). Here flat priors for the connectives and noise levels are used for each new situation.

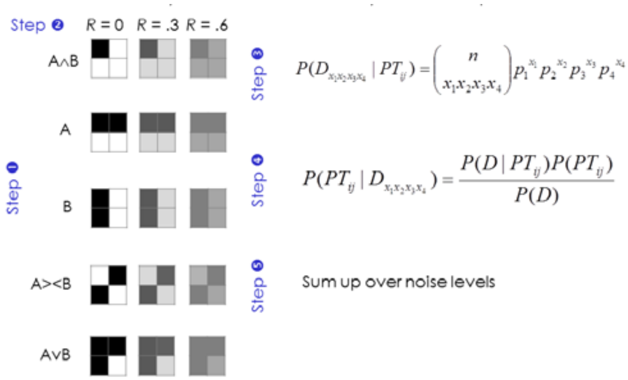

Figure 1: Sketch of the model of only five logical propositions and three noise levels (cf. text for details).

For the novel noisy-logical representation one can calculate the posterior probabilities for each probability table by combining some standard statistics. Given observed data about the co-occurrence of *A* and *B* (a 2 by 2 contingency matrix), one can calculate (Step 3) the likelihood of the data given a probability table, *P*(*D* | *A*o*B*, *R*), by using the multinomial distribution, which here determines for each table of four probabilities how likely it produces the observed four frequencies. In Step 4, Bayes' theorem is used to transform the likelihood *P*(*D*| *A*o*B*, *R*) into a posterior *P*(*A* o *B*, *R* | *D*). In a final step, one sums up the probabilities of a connector over all noise levels (here we modeled 11 equidistant levels from R = 0 to 1). We obtain the requested posterior *pattern* probability, $P_P(A\text{o}B \mid D)$, clearly differing from the analogous *extensional* probability, $P_E(A\text{o}B|D)$ (frequency of positive cases, divided by all cases).

BL and the Conjunctions Fallacy Debate

One of the most heated and philosophically interesting psychological debates concerns the apparent inability of people to understand that conjunctions (for instance, "Linda is a Bank teller and a feminist") can never be (extensionally) more probable than their conjuncts (e.g., "Linda is a bank teller")—even for apparent feminists. This phenomenon has been called "conjunction fallacy" and first has been explained by the representativeness heuristic (Kahneman & Tversky, 1982). This heuristic, however, has been criticized as being formulated too imprecise (Gigerenzer, 1996; cf. Nilsson, Juslin, & Olsson, 2008).

There have been several other classes of explanations of CFs. One focusses on possible misunderstandings. "*A* and *B*" may actually be understood as "*A* or *B*" or to "if *A* then *B*" (Mellers, Hertwig, & Kahneman, 2001; Hertwig, Benz & Krauss, 2008). Moreover, "*A*" may be interpreted as "*A* but not *B*" instead of "A, whether *B* or not *B*" (Kahneman & Tversky, 1983; Hilton, 1995; cf. Sides, Osherson, Bonini, & Viale, 2002; Wedell & Moro, 2008). A second class of explanations considers different ways in which probabilities are introduced and how the probability question is posed. It has been shown that frequency presentations (Fiedler, 1988; Gigerenzer, 1996), rating formats (Sloman, Over, Slovak, & Stibel, 2003), and clear set inclusions (Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 1999; Sloman et al., 2003) often substantially reduced the portion of CFs. Although these factors often do play a role, BL under certain conditions (if one is concerned with alternative hypotheses about whole situations) has predicted CFs even when all these factors apply simultaneously (von Sydow, 2011a, b). A third class of explanations specifies quantitative conditions of CFs. Most prominently, it has been suggested that the requested probability $P(A \wedge B \mid D)$ is replaced by the inverse probability $P(D \mid A \wedge B)$; cf. Wolford, 1991; Fisk & Slattery, 2005), or by a measure of support, like $P(A \wedge B \mid D)$ - $P(A \wedge B)$ (support theory, cf. Sides, et al., 2002; Lagnado & Shanks, 2002; cf. Tentori, Bonini, & Osherson, 2004; Crupi, Fitelson, & Tentori, 2008), or several other measures, like signed summation, averaging, quantum logic, or rescaling (see Wedell & Moro, 2008; von Sydow, 2009).

BL provides a rational quantitative account of frequency-based a particular class of conjunction fallacies and made several novel predictions that cannot be explained by the previous models (von Sydow, 2011). One important aspect has been the generalization of the idea of CFs into a system of logical inclusion 'fallacies' (von Sydow, 2009).

Experiment: Trial-By-Trial Induction of Logical Relationships

The primary goal of the reported experiment is to test aspects of the postulated system of frequency-based logical inclusion 'fallacies' in a trial-by-trial way. Whereas confirmatory results for this system have already been achieved, even using trial-by-trial *presentation* of items (von Sydow, 2011b; cf. Lagnado et al., 2001), we here additionally investigate trial-by-trial *assessment of the dependent variable*: the selection of the most probable hypotheses after each new observation. To the best knowledge of the authors, this has never been investigated before in the CF debate.

A supplementary goal is to assess whether putting evidence into language in the course of trials may have an additional top-down effect on the successive evaluation of evidence. Here the *ways* how one obtains a final (fixed) pattern of evidence are varied, so that this may affect the predicted propositional representations. In one condition the finally predicted hypothesis is expected to appear most probable all along (homogeneous condition) and in another condition different hypotheses are predicted to appear more probable throughout the first learning trials (heterogeneous condition). In its current formulation BL, as a model of data-based induction, would not be able to account for such top-down effects. This is the case although BL goes beyond naïve probability, and leaves room also for subjective priors. As we think there are top-down effects for instance of categorization (Hagmayer, Meder, von Sydow, & Waldmann, 2011) or causal coherence (von Sydow, Hagmayer, Meder, & Waldmann, 2010), we think there may well be top-down-effects of mere verbalization. In this experiment, however, participants are provided with summary statistics, excluding memory effects. In such settings, also intended as base-line for future experiments, no such additional top-down effects are expected.

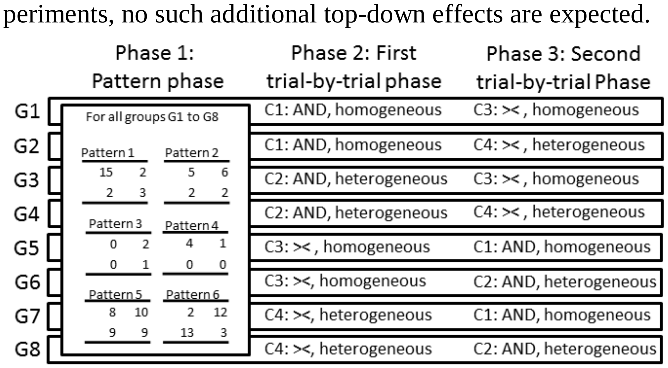

Figure 2: Design (see main text for details).

The design involves three phases. All phases involve a selection of the most probable logical hypothesis given some evidence. In *Phase 1*, participants in all conditions are randomly presented with six patterns of evidence, each referring to a different situation (Fig. 2, Phase 1). First, this phase should replicate previous generalizations of BL (von Sydow, 2009, 2011b). Secondly, it investigates whether participants grasp the intended meaning of logical terms, and,

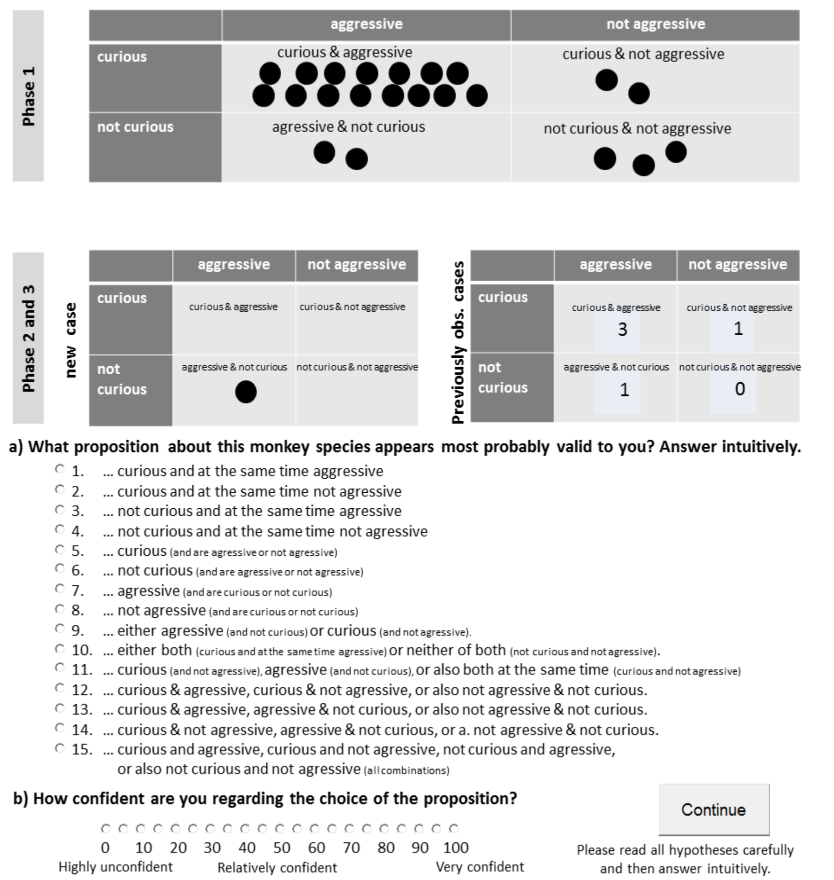

Figure 3: Format of the frequency presentation in Phase 1 (patterns) and in Phase 2 and 3 (trial-by-trial) and formulation of hypotheses.

thirdly, it excludes a deterministic understanding of the rules in the next phases by inducing a non-deterministic noise-prior (especially for few observed cases, priors may well affect the results).

Phase 2 and 3 are both trial-by-trial judgment tasks. BL predicts that various hypotheses should be selected to be most probable, each from an extensional perspective involving several logical inclusion fallacies. The sequences should end up either in an AND hypothesis (C1, C2) or an EITHER-OR hypothesis (C3, C4). Both hypothesis are extensionally less probable than the OR hypothesis or the tautology. Additionally, the order in which data is presented differs, investigating whether verbalization throughout learning affects the verbalization of identical final patterns (the probability judgments). As sketched, either a homogeneous condition (C1, C3) or a heterogeneous condition (C2, C4) is used. Finally, Phase 2 and 3 are identical, in order to assess whether the previous learning phase had an effect (as, e.g., suggested by support theory) and to find out whether participants increasingly make either extensional or BL selections.

Material

130 participants of the University of Göttingen participated in the experiment. The participants were told about newly

discovered species of apes on a lonely island. They were in the role of ethologist concerned with statements the animals of a species are curious or not (here *A*) and whether they are aggressive or not (here *B*), as well as judging the relation of these properties.

In Phase 1 participants were concerned with six species of apes in randomized order. For each species they were shown a photo of an ape (e.g., “P. calvus”) with a text “The animals of this species are…”, leading to the main instructions (Fig. 3) and a contingency table summing up the observed features combinations (cf. Fig. 2, 3). For each species one had to select the most probable logical hypothesis and one had to provide a confidence rating (Fig. 3).

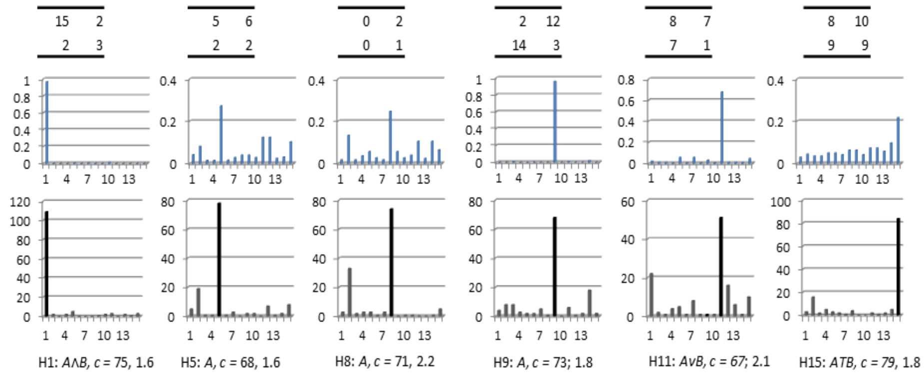

Figure 4: Six patterns of Phase 1 (first row), BL’s pattern probabilities (second row) and the frequency of hypotheses (cf. Fig. 3) selected (third row).

Phase 2 and 3 were concerned with trial-by-trial learning. Participants were randomly assigned to the eight conditions. Single events were symbolized by a circle flying to a place in the contingency table (Fig. 3, Phase 2/3, left table), followed by an update of a summary table (right table). Of the 18 trials the first nine are presented in Figure 5 and 6.

In all probability judgment tasks the formulations of the hypotheses were carefully chosen to rule out the plausible misunderstandings discussed in the CF debate. For instance, the conjunctions were formulated as “*A* and *at the same time B*” and the single conjuncts (the affirmations or negations) as “*A* (and are *B* or not *B*)” (Fig. 3).

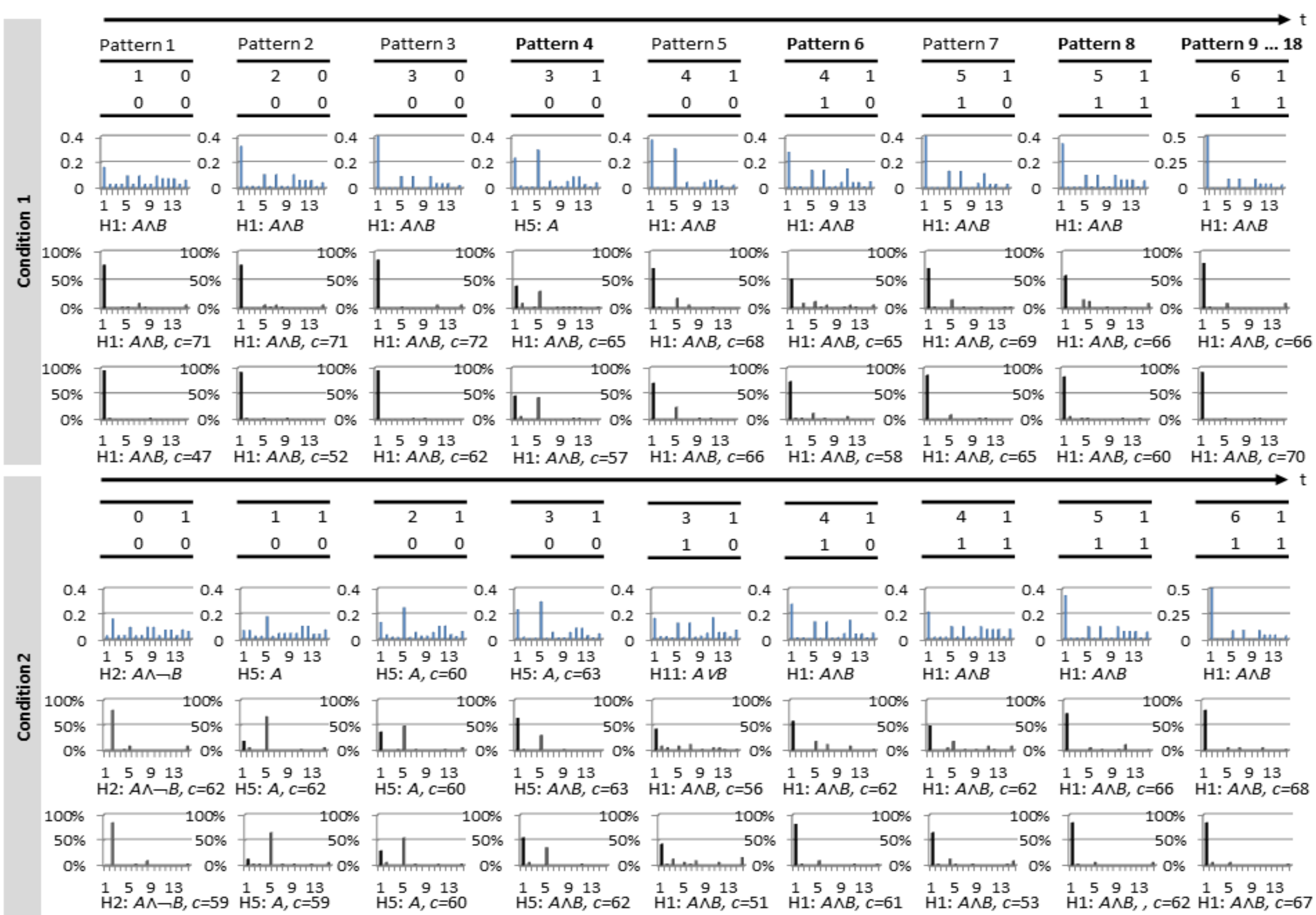

Figure 5: Patterns, predictions and results for Condition 1 and 2 (Phase 2 and 3). Within each condition, Row 1 shows the first nine shown patterns (Fig. 3, Phase 2 and 3, right). Row 2 depicts BL’s pattern probabilities for 15 hypotheses (cf. Fig. 3). Row 3 and 4 show the portion of hypotheses actually selected to be most probable (in Phase 2 and 3).

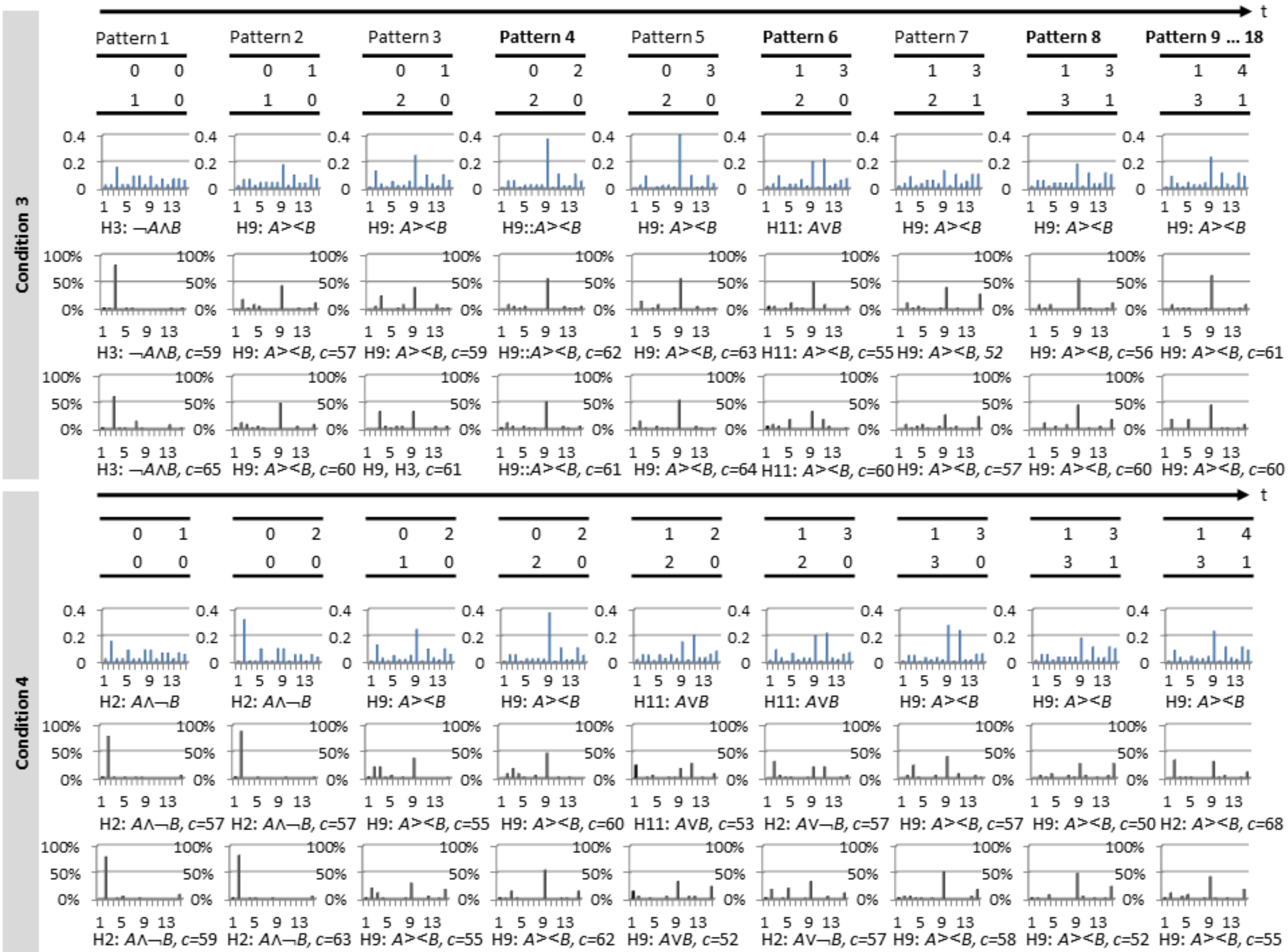

Figure 6: Patterns, predictions and results for Condition 2 and 3 (cf. Fig. 4 and main text for details).

Results

Figure 4 shows the presented data patterns, the predicted pattern probabilities (BL), and the empirically found frequency of selected logical hypotheses for the six shown species of apes. Participants for each pattern actually selected the hypothesis that had the highest pattern probability, $P_P(A \text{ o } B)$; from left to right: H1 (*A* and *B*); H5 (*A*); H8 (not-*B*); H9 (either *A* or *B*), H11 (*A* or *B* or both), H15 (everything is possible). If one extended other theories so that they may predict these connectives, one would presumably not to be able to explain the data (cf. von Sydow, 2010). For instance, the interesting support theory would make predictions for Pattern 6 [8, 10, 9, 9] based on the five other patterns (resulting in sum in [30, 29, 25, 10]). The highest support is suggested for the *d*-cell (H4) which is actually found only rarely. The strongest deviation from BL is observed in Pattern 5 were participants did *not only* select H11 but also H1. But this needs not to refer to an alternative strategy, but perhaps – and without elaborating this here – with a noise prior excluding deterministic patterns and causing the actual outcome (cf. von Sydow, 2011b).

With regard to Phase 2 and 3, Figures 5 and 6 show for all conditions the presented data sequence, the resulting BL probabilities, and the actually observed frequencies of the selections of the most probable hypotheses. Even for the low trial numbers 1 to 9, reported here, the main selections are generally surprisingly in line with the pattern probabilities (presented without any fitting).

There were only small deviations. For instance, in Condition 1 only in Pattern 4 the predicted mode of answers (H5: *A*) differed from the observed one (H1: $A \wedge B$). However H1 has actually the second highest pattern probability and there may again have been a plausible influence of noise priors resulting from Phase 1 (lowering $P(R = 0)$), which would actually increase $P_P(H1)$. This would likewise be coherent with Pattern 6 [4 1 0 0], were a surprisingly clear majority choose the AND-hypothesis (H1) and the extensional answer would be the *A*-hypothesis (H5).

The patterns that were kept identical in the corresponding homogeneous and heterogeneous conditions (the bold printed Patterns number 4, 6, 8, and 9) mostly corroborated the same results, suggesting that if memory effects are ruled out (as done here), no or only small effects of homogeneous versus heterogeneous conditions are obtained. Furthermore, as predicted based on BL, the results were more pronounced for the conjunction conditions than for the exclusive disjunctions. Finally, the outcomes of Phase 2 and 3 did not differ much (or the results for BL even improve over time).

The confidence ratings varied less clearly than expected. One reason may be that this measure reflects not only, for instance, P_H(H most probable)/P_P(H second most probable), but a general belief in a system of answers corresponding to BL or extensional probabilities. Furthermore, the ratings, averaged over all participants, may not be diagnostic, since they include ratings of unpredicted hypotheses (particularly relevant in C3 and C4). However, at least in the second trial-by-trial phase (Phase 3) participant's confidence ratings roughly corresponded to predictions derivable from BL: In C1 confidence increases from Pattern 1 to 3. In Pattern 9 the confidence is higher than in all previous patterns (despite more outliers). For Condition 3 and 4 the ratings show less differences, as is understandable based on pattern probabilities. Nonetheless, if one additionally takes a look at the next repeated nine trials, not reported here, Trial 18, for instance, confirmed a high confidence, leading to a median of 80 in C3 and 70 in C4. Hence, also the confidence ratings, at least in Phase 3, strongly reflect changes coherent with BL.

Discussion

The results show correspondence with the predictions of BL also in trial-by-trial probability judgment tasks. Although other models of the CF have not been extended to all other connectives, it seems implausible that they could account for the findings (cf. von Sydow, 2009, 2011a). Without being able to discuss this here, some deviations (but clearly not all findings) may be coherent with a model that I have previously called pattern support, combining the pattern idea of BL with the idea of support. Overall, however, the results provide additional evidence for the predicted class of frequency-based CFs and for BL as a (computational level) psychological model for noisy-logical relationships.

Furthermore, as expected the results show no (or only a small) top-down effects of verbalization of hypotheses about the same situation (homogeneous vs. heterogeneous conditions). In the future it will be interesting to investigate identical settings without memory hooks (without summary statistics in Phase 2 and 3). Then verbalization may well effect represented exemplars (cf. von Sydow, 2011b). A further line of future research should be to investigate the role of noise priors on the selection of hypotheses.

Acknowledgments

The first author has been the main author of this paper. We thank H. Wilke and J. Frisch for support. The first author is grateful for inspiring discussions at the Universities of Göttingen (M. R. Waldmann, R. Mayrhofer, Y. Hagmayer) and Heidelberg (group of K. Fiedler). This work was supported by the grant Sy 111/2-1 to M. von Sydow from the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program *New Frameworks of Rationality* (SPP 1516).

References

Chater, N., Tenenbaum, J., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Science, 10*, 287-291.

Crupi, V., Fitelson, B, & Tentori, K. (2008). Probability, Confirmation, and the Conjunction Fallacy – Theoretical Note. *Thinking and Reasoning, 14(2)*, 182 - 199.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research, 50*, 123-129.

Gigerenzer, G. (1996). On narrow norms and vague heuristics. *Psychological Review, 103*, 592-596.

Hagmayer, Y., Meder, B., von Sydow, M., & Waldmann, M. R. (2011). Category Transfer in Sequential Causal Learning: The Unbroken Mechanism Hypothesis. *Cognitive Science, 35*, 842-873.

Hertwig, R., Benz, B., & Krauss, B. S. (2008). The conjunction fallacy and the many meanings of and. *Cognition, 108*, 740-753.

Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment, *Psych. Bulletin, 118*, 248-271.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S. & Caverni, J.-P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106*, 62-88.

Kruschke, J. K. (2008). Bayesian approaches to associative learning: From passive to active learning. *Learning & Behavior, 36*, 210-226.

Lagnado, D. A. & Shanks, D. R. (2002). Probability judgment in hierarchical learning: A conflict between predictiveness and coherence. *Cognition, 93*, 81-112.

Mellers, B. A., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science, 12*, 269-275.

Nilsson, H., Juslin, P., & Olsson, H. (2008). Exemplars in the mist: The cognitive substrate of the representativeness heuristic. *Scandinavian Journal of Psychology, 49*, 201-212.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality. The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Schurz, G. (2005). Non-monotonic reasoning from an evolutionary viewpoint: ontic, logical and cognitive foundations. *Synthese, 146*, 37–51.

Sides, A., Osherson, D., Bonini N., & Viale, R. (2002). On the reality of the conjunction fallacy. *Memory and Cognition, 30*, 191-198.

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions. *Organizational Behavior and Human Processes, 91*, 296-309.

Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: a misunderstanding about conjunction? *Cognitive Science, 28*, 467-477.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293-315.

von Sydow, M. (2009). On a General Bayesian Pattern Logic of Frequency-Based Logical Inclusion Fallacies. *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society* (pp. 248-253). Austin, TX: Cognitive Science Society.

von Sydow, M. (2011 a). The Bayesian Logic of Frequency-Based Conjunction Fallacies. *Journal of Mathematical Psychology, 55(2)*, 119-139.

von Sydow, Momme (2011 b). Logical Inclusion Fallacies - Transfer of Logical Patterns and Noise (pp. 1-6). In: Kokinov, B., Karmiloff-Smith, A., Nersessian, N. J. (eds.). *European Perspectives on Cognitive Science*. New Bulgarian University Press.

von Sydow, M., Hagmayer, Y., Meder, B. & Waldmann, M. (2010). How Causal Reasoning Can Bias Empirical Evidence. *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society* (pp. 2087-2092). Austin, TX: Cognitive Science Society.

Wedell, D. H., & Moro, R. (2008). Testing boundary conditions for the conjunction fallacy: Effects of response mode, conceptual focus, and problem type. *Cognition, 107*, 105-136.