Pattern Probabilities for Non-Dichotomous Events: A New Rational Contribution to the Conjunction Fallacy Debate

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

Abstract

This paper analyzes probability judgments about properties that can take multiple values (i.e., monadic polytomous events). It extends previous work on pattern-based deviations from standard (extensional) probabilities. Pattern-probabilities are formalized in Bayesian Logic (BL) and should be applicable when testing the overall adequacy of alternative logical hypotheses while allowing for exceptions. BL systematically predicts 'conjunction fallacies' (CFs) and, more generally, 'inclusion fallacies' (IFs), when a subset is deemed more probable than its superset. BL formalizes a fit between data and explanatory noisy-logical patterns and was supported in previous experiments on *dyadic* logical connectives with two *dichotomous* events. Here BL is extended to *monadic* prediction with *several* subclasses. BL may for instance predict $P_{pattern}(A) > P_{pattern}(\text{non-}A)$ even though $f(A) < f(\text{non-}A)$, given that non-*A* has more subclasses than *A*. Two experiments using material from the Linda paradigm corroborate a pattern approach and rule out confirmation as an alternative explanation.

Keywords: probability judgments; biases; conjunction fallacy; inclusion fallacy; inductive Bayesian logics; predication; strong sampling; categories; Lockean Thesis

We cannot dispense with the idea of the rationality or irrationality of judgments, even though defining their rationality may often be difficult. The most established and important norms of rationality are presumably basic norms of formal logics and probability theory. As early as over 2000 years ago, Aristotle's Organon provided a systematization of logics and also, albeit less refined, of inductive belief. The history of philosophy and mathematics in recent centuries led to canonical formulations of formal logics (Frege, Russell, Whitehead, Wittgenstein) and probability theory (e.g., Kolmogorov). Although modifications and more specialized systems continue to be developed in philosophy, mathematics and computer science (e.g., modal logics, multivalued logics, non-monotonic logics, belief functions), psychology continues to focus almost exclusively on these basic norms. Their domain-general application molded Kahneman and Tversky's heuristic-and-bias approach, taking these norms as given and concluding that *homo sapiens sapiens* is ultimately irrational, even with regard to most basic laws of rationality (cf. Fiedler & von Sydow, 2005, evaluation of this fertile research program).

Kahneman and Tversky studied for instance a scenario wherein Linda, *L*, is described by a story in a way that seems suggestive of her being a feminist. Participants were then asked whether *P*(*L* is a bank teller) or *P*(*L* is a bank teller and a feminist) is higher (Kahneman & Tversky, 1982). Most participants judged $P(B) < P(B \,\&\, F)$, thus committing a conjunction fallacy (CF), since the (standard) probability of a logical conjunction can never be higher than of one of its conjuncts.—This bias-and-heuristic approach has been criticized, fiercely but eloquently, by Gigerenzer (1996). He criticized not only the alternative heuristics (in the above example, 'representativeness') as "one-word theories" almost void of any explanatory value, but also the "content-blind" application of "narrow norms" without accounting for their ecological context. Although my approach clearly differs from Gigerenzer's explanation of the CF (von Sydow, 2011), I think he is right in stressing that rational norms are essentially contextual. However, I agree with Kahneman (1996) that criticizing 'narrow norms' risks the danger of normative agnosticism.

This paper addresses the extensionality inherent in standard norms of logics and probability theory, by investigating probability judgments about several mutually exclusive classes. Standard probability is extensional since the probability of a monadic hypothesis (*X* are *A*) is defined by the relative frequency of confirmatory elements (its extension) relative to all elements in a universe of discourse. Extensionality does not consider the number of subclasses (an essential aspect of *in*tension), but rather the number of observed cases in a class. We here apply Bayesian Logic (von Sydow, 2011) to monadic predicates based on several explicitly represented subclasses, where the intension of the class matters. It also builds on the renaissance of Bayesian approaches in cognitive science (Oaksford & Chater, 2007). It also builds on the idea of strong sampling (Tenenbaum & Griffiths, 2001; cf. Navarro, Dry, & Lee, 2012), but BL can even assign higher probabilities to more specific hypotheses if there are known exceptions (von Sydow, 2011).

In philosophy the Lockean Thesis refers to the controversial idea that one can assign probabilities to logical propositions (Foley, 2009). The non-standard approach of BL may allow for rescuing common sense assumptions and a kind of Lockean Thesis—perhaps better called 'Bayesian thesis' here.

Truth Table Logics and Probability Theory – Two Narrow Norms for Predication?

Problem of sample size Given one has no prior knowledge about a group of animals, *X*, and has either observed one or one hundred confirming cases to be black, *B*, standard logics does not distinguish between these situations, and even relative frequencies assign a probability of 1 in either case. However, the use of probabilities about probabilities seems more reasonable.

Problem of exceptions We believe "ravens are black", even though we know that white ravens do exist. Should exceptions falsify nearly all our general predications? To use standard probability instead of logics seems to resolve this problem, since the probability $P(B|X)$ remains high.

Problem of inclusion However, even if we have observed *Xs* being *B* a hundred times, and only a single case 'X is non-*B*', $P(B|X)$ would remain smaller than $P(B \vee non\text{-}B|X)$. Taking probability as the criterion for adequate predication would in uncertain situations still force us always to prefer tautologies (such as "Ravens are black or not black") over any other, more informative hypotheses.

Hence, if an adequacy criterion of predication should describe the real contingent patterns found in the world, neither standard logics nor probability could serve as a suitable adequacy criterion (von Sydow, 2011).

Bayesian Logic and One-Dimensional Pattern Probabilities of Several Subclasses

Dyadic Bayesian Logic Von Sydow (2011, 2014) has argued that when one is concerned with alternative hypotheses about the interaction of two dichotomous attributes (as described by dyadic logic), and if people are concerned with predications describing an overall situation, pattern probabilities, specified by Bayesian Logic (BL), seems to provide a suitable alternative adequacy criterion. BL specifies ideal logical explanatory patterns, adds noise, and then calculates posterior probabilities for these ideal noisy-logical explanations given some data. BL allows for assigning higher probabilities to more specific hypothesis even in the light of exceptions. A corresponding class of inclusion 'fallacies' is thus interpreted as rational. The Linda scenario, for instance, with some plausible auxiliary assumptions may yield $P_P(A) > P_P(A \wedge B) > P_P(B)$. BL has been supported in several studies using precise frequency input and investigating a system of inclusion fallacies (e.g., von Sydow & Fiedler, 2013; von Sydow, 2014).

One-dimensional (monadic) predication with several subclasses Most inclusion fallacy studies have focused on dyadic connectives such as "AND, EITHER OR, etc.". In contrast, here we concentrate on probability judgments about predicates concerning a single dimension only. In logical parlance, we are concerned with monadic connectives, such as: "*X* are *A*" (Affirmation), "*X* are non-*A*" (Negation), "*X* are *A* or Non-*A*" (Tautology) (cf. von Sydow, 2014). Here we focus on monadic predication involving *more than two* represented qualitative classes. For example, consider the (main) jobs of the graduates of a school: Individually the outcomes (bank teller *B*, translator *C*, etc.) should be mutually exclusive. For "pupils from the Linda schools become bank tellers or translators", $B \vee T$, standard probability judgments do not allow for $P(B \vee T) < P(B)$, whereas BL may, depending on data and assumed noise, rationally predict such IFs. Another aspect of BL is that it might predict $P(A) > P(Non\text{-}A)$ even if $f(A) < f(\text{non-}A)$. Such findings exclude many but not all alternative models (Tentori et al., 2013; cf. von Sydow, 2011, Exp. 2).

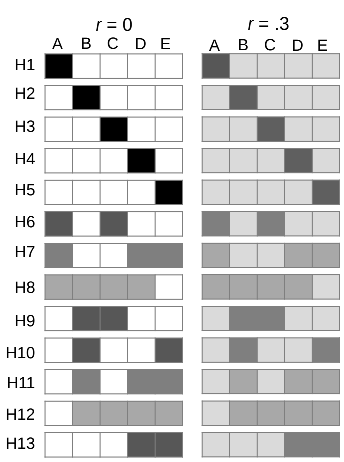


Figure 1: Thirteen hypotheses about five true/false classes as ideal-noisy patterns (or probability tables) at two levels of potential noise r = 0; .3.

Monadic BL with several subclasses Subsequently BL is formalized for monadic predication with *c* qualitatvive subclasses. However, we will focus on $c = 5$: *A*, *B*, *C*, *D*, *E* (cf. the experiments for examples). Hypothesis '*Xs* are *A*' [*A*] is thus nested in the hypotheses '*Xs* are *A or B*' [*AB*], and in *AC*, or *ABC*. Therefore, standard (extensional) probabilities require $P_E(A) \leq P_E(AB) \leq P_E(ABC) \leq P_E(ABCD)$.

Step 0 only describes the input which should then be compared to the ideal patterns specified in Steps 1 and 2. BL may take frequencies as well as beliefs (about extensional probabilities of cells) as input. BL assumes a standard belief-update of cell probabilities for (pre-categorized) alternative *c* classes. A multinomial update results in beliefs modelled by a Dirichlet distribution, with priors and event frequencies for each class. Likewise, for instance memory-based, distortions of actual frequencies (which are not part of our model) may still be captured by the distribution, simply by using subjective rather than objective frequencies. The Dirichlet distribution can be used to code beliefs in terms of frequencies. In sum, the model takes the *subjective* frequencies or beliefs in the cell frequencies (or cell probabilities) as input. This, however, does not resolve the problem of inclusion. The Dirichlet distribution flexibly formalizes actual belief (coherent with *any* subjective joint frequency distribution), yet it does not specify the belief in ideal explanatory patterns given such data.

Step 1 treats hypotheses such as *AC* or *ABCD* as ideal explanatory patterns that may have generated the data. Figure 1 shows 13 such patterns/hypotheses (e.g., H8 ABCD, or H9 AC). The left panel for no noise (r = 0) models patterns without exception tolerance. A single counterexample still falsifies a hypothesis (e.g., a single *D* observation falsifies H9). For ideal explanatory patterns we assume equiprobability for confirming classes. The ideal cell probability for confirmatory cells is $p_{conf} = 1/c_{conf}$, with c_{conf}, being the number of confirmatory cells in a pattern (e.g., H7: p_{conf} = .33; cf. Tenenbaum & Griffiths, 2001). This replaces (monadic) truth tables by ideal probability tables. The cell-probabilities are represented in Figure 2 by shades of grayscale (white = 0; black = 1).

In *Step 2* further possible equidistant levels of noise *r* are added in the interval [0, 1] (in Panel 2, e.g., r = .3). The disconfirmatory classes of a pattern here get ideal cell probabilities above zero: $p_{dis} = r/c_{all}$. Correspondingly, the confirmatory cell probabilities have to be reduced: $p_{conf} = 1/c_{con} - r(1/c_{con} - 1/c_{all})$. This results in ideal noisy-logical probability tables, PTs, taken as possible (monadic) explanations given the data. Based on these patterns the following steps proceed in a fairly standard way.

Step 3 calculates the likelihood of the data δ given the ideal explanatory hypotheses. δ, in our polytomous monadic case, consists of a vector of length *c*, $\delta_1,\ldots,\delta_c$, (cf. Step 0). For a particular probability table PT (Step 2, 3) referring to a hypothesis *j* and a noise level *i*, we calculate the likelihood using the multinomial distribution. In our example (*c* = 5):

$$P(\delta|PT_{i,j}) = \binom{n}{\delta_1\delta_2\delta_3\delta_4\delta_5} {p_1}^{\delta_1}{p_2}^{\delta_2}{p_3}^{\delta_3}{p_4}^{\delta_4}{p_5}^{\delta_5}$$

In *Step 4*, Bayes' theorem is used to turn the likelihoods into posterior probabilities:

$$P(PT_{i,j}|\delta) = \frac{P(\delta|PT_{i,j})P(PT_{i,j})}{P(\delta)}$$

Here $P(PT)$ represents the prior for a PT (in line with the subjective frequencies defining our subjective belief distribution; cf. Step 0). $P(\delta)$ is a normalizing constant (calculated by summing up all likelihoods weighted by their priors).

In *Step 5*, the probability of a particular hypothesis *i* is determined by summing up over the PTs with its different noise levels *j*. (Here no further steps or weightings are needed; cf. von Sydow 2011.) The one-dimensional pattern-probabilities differ considerably from standard extensional probabilities and from other measures suggested to model IFs. In the following, we start investigating monadic polytomous BL and the predicted class of IFs empirically.

Experiments 1 and 2

Here I report only considerable parts of two simple experiments concerning probability judgments single polytomous attribute dimensions. (monadic predication).

The experiments investigate (monadic) inclusion fallacies (IFs) since the hypotheses are partially nested. The classes in this dimension, *A*, *B*, *C*, *D*, *E*, are framed as mutually exclusive, and one prediction of BL is that people judge $P(A) > P(\text{non-}A)$ with $P(\text{non-}A) = P(B) + P(C) + P(D) + P(E)$ even if $f(A) < f(\text{non-}A)$. The predictions for pattern probabilities, P_P, clearly differ from those based on standard extensional probabilities (relative frequencies), P_E.

The predictions of BL also differ, for instance, from a confirmation account of IFs that assumes that people misinterpret probability as confirmation, which should result in selecting the most strongly confirmed hypothesis when asked for the most probable one (cf. von Sydow, 2011).

Although we focus on *monadic* IFs, we use attributes reminiscent of the original Linda task: In Experiment 1 we are concerned with job descriptions, in Experiment 2 with political attitudes. Additionally, in Experiment 1 we use either ordinary language "OR" or "AND" to sum several subclasses. Although the latter seems logically incorrect, we predicted that in a context of mutually exclusive events the meaning of these 'connectives' would not differ. Furthermore, in both experiments the order of labels of the classes was counterbalanced to control for content effects. Finally, in Phase 1 we investigated single samples of data and in Phase 2 a timed series of three samples to test for contrasting predictions of priors vs. confirmation.

Method

Materials and procedures Both experiments concern thirteen alternative statements about the graduates of different schools (Linda school, Humboldt school, Goethe school etc.). The experiments were run on a computer.

On the *introduction pages*, participants were told that they would receive sample information about the distribution of graduates, and they were given two examples for sample distributions. In Experiment 1, the attribute-dimension concerned five classes of graduates' main jobs (*A* translator, *B* bank employee, *C* artist, *D* teacher, *E* physician). Experiment 2 concerned five classes of main political attitude (*A* conservative, *B* social, *C* liberal, *D* ecological, *E* feminist), explicitly framed as alternative outcomes of a test that assigns people to *one* political group only. With regard to procedure and samples the experiments were identical.

In *Phase 1*, fifteen schools with counterbalanced names were shown in random order. For each school they saw one sample of graduates: a table with category names and frequencies (numbers) showed which jobs the sample had chosen. Apart from the category order *A, B, C, D, E,* a second condition involved the label orders *E, D, B, A, C* (experimental factor 1 in both Experiments) to control for content effects/prior knowledge: Ecological analysis suggested that if exogenous prior probabilities were considered, *D* and *E* should subjectively be expected the most probable. Likewise for Phase 1 and the first label order summing up all samples involved $P_E(A)$ = 13%, $P_E(B)$ = 24%, $P_E(C)$ = 10%, $P_E(D)$ = 27%, $P_E(E)$ = 25% (since the frequencies are kept constant in the second condition, other labels had high/low values). BL should only be weakly influenced by such priors, since a strong transfer between different schools seems implausible, particularly if the present samples are large enough that it is clear they come from different populations. However, a confirmation approach may rely on these differences (cf. also Phase 2).

Participants were asked: "Which statement appears most probably to be valid to you? [...] Answer intuitively." The experiments concerned: "Pupils of this school…". In Experiment 1, the hypotheses concerned whether the pupils of the school become for example "S1 translators", or "S8 translators, bank tellers, artists or teachers". Note that S1 refers to a subset of S8. In Experiment 2, hypotheses likewise concerned nested attributes such as: "S1 …conservative" or "S8 …conservative, social, liberal or feminist".

In addition to the logically correct OR-formulations, Experiment 1 used AND-formulations (experimental factor 2): "S8 translators, bank tellers, artists and teachers". If AND were here interpreted as logical conjunction, the intersection would be empty, but for alternative classes AND was expected to be used for adding as well.[1]

Apart from the logical formulations and the varied labels, the thirteen hypotheses in both experiments referred to the

[1] Experiment 2 did not pursue this issue, since the AND-interpretation as logical conjunction may become more plausible due to a non-exclusive interpretation of classes such as "conservative" and "feminist".

same classes; as shown in Figure 2 in Phase 1 this is: S1 A; S2 B; S3 C; S4 D; S5 E; S6 AC; S7 ADE; S8 ABCD; S9 BC; S10 BE; S11 BDE; S12 BCDE; S13 DE.

Phase 2 of the experiments concerned ten further schools in random order. Participants now for each school were successively provided with samples from three surveys of differing dates, one twenty years earlier, another ten, and then with a recent survey. It was mentioned that the surveys might differ in size. Participants should select the hypothesis that appeared most probably valid with regard to the *present* school. The hypotheses in Phase 2 (cf. Figure 2, Panel 2, cf. the abscissa of the graphs), apart from the label and the formulation factor, did not differ in the experiments.

Participants In Experiments 1 and 2, 57 and 36 participants from the University of Heidelberg took part voluntarily, in exchange for course credits or money (8 Euros per hour). They were assigned randomly to the four conditions of Experiment 1 and the two conditions of Experiment 2.

Results

Figure 2 shows the proportion of selected hypotheses judged to be most probable in the six selected schools *S* investigated in Phase 1 (with a single sample per school), and the six selected schools in Phase 2 (three successive samples). We selected most informative examples (including the cases with most deviation between the two experiments: S3, S5). For simplicity, we conflate over the experimental conditions of label order and formulation type. But out of the 25 data patterns in overall six conditions (hence 150 charts), the main prediction of BL deviated from the predicted modal selection in two cases only. As predicted, it does not seem to matter whether the addition of classes was expressed by an OR or an AND in ordinary language. The high fit of the predictions of BL, however, suggests that people understood the relationship between language and logical meaning.

Conflated over conditions, the modal selection in all 25 schools (both phases) corresponded to the hypothesis predicted by BL. The mean correlation in the schools between selections and deterministic prediction were: Exp. 1, .98; Exp. 2, .96. In all schools, *S*, (both Experiments) the same modal values resulted—despite different content.

Subsequently we discuss results for the example schools in Phase 1 and in Phase 2. In school *S1*, $P_E(A)$ is less extensionally probable than its negation $P_E(B)+P_E(C)+P_E(D)+P_E(E)$. Nevertheless the vast majority of participants, in line with BL, judged hypothesis *A* to be more probable than hypothesis '*BCDE'* (with $P_P(A) < P_P(BCDE)$). This involves systematic inclusion fallacies (IFs) predicted by BL (since the hypotheses *AC*, *ABE*, *ABCD* have higher extensional probabilities). The difference between the experiments or corresponding two label conditions seemed week. Additionally, for another school—not reported here—a exactly reverse order of frequencies led to highly analogous results.

S2 excludes that people always select the modal answer (*B*), since people, as predicted by BL, mostly selected the *BDE* hypothesis. For a confirmation account, prior cell beliefs should have led to more varying selections in the two label order conditions. Likewise, confirmation cannot explain the results if one assumes flat prior beliefs, since it would be highest for the modal-frequency hypothesis (*B*).

S3 corroborates that people select the inclusive *BCDE* hypothesis if predicted, although confirmation (for flat priors) would predict selection of *E* instead of *BCDE*.

S4 excludes a possible simpler approximation of pattern probabilities by a 'ratio heuristic'. The break-point between attributed and non-attributed classes might have been the largest ratio between classes. However, despite 6/1 > 29/6, the results corroborate the prediction of $P_P(DE) > P_P(BDE)$.

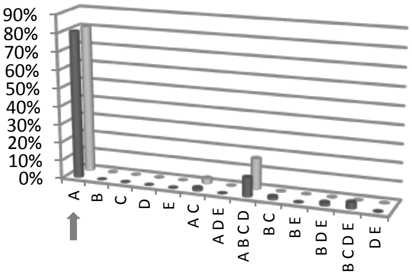
S5 investigates a situation in which BL's prediction DE is less clear (see the second, minor prediction *D*). Participants in Exp. 2 (political attitude) selected more frequently the more specific hypothesis (*D*) over the main prediction (*DE*) – perhaps due to chance, or to greater noise-tolerance in the political scenario (cf. S3, S11), which could be accounted for within BL.—*S6* led to the predicted dominant inclusive selection *ADE*, not to the hypothesis referring to the modal frequency *A*. However, in at least one of the labeling conditions, *A* may have had lower prior expectations and hence a higher confirmation than classes *D* or *E* (and hence than the overall hypothesis *ADE*). Even if again we alternatively assume flat priors for the five classes, the standard extensional conceptualization of confirmation predicts the selection of the most specific hypothesis, here *A*.

Panel 2 of Figure 2 shows the results for *Phase 2* and which hypotheses are thought to hold most probable for a present school (*t3*), after showing two previous samples from earlier years (*t1*, *t2*). Although it is not clear to what *extent* previous data should be used for the judgment about *t3*, an inductive transfer parameter may rationally model the degree to which old data should be used in new situations. However, the *directions* of prior effects are clear and they can be contrasted with predictions of confirmation. For BL the influence of a prior should be mainly effective if one has small data samples in *t3*, otherwise the data may suggest that earlier samples come from different populations and should be ignored.—In *S7* and *S8* the small sample in t3 according to BL does not favor *DE* very strongly over hypothesis *E*. As predicted, the selections (in both Experiments) varied in line with the t1 and t2 priors, favoring either *E* (S7) or *DE* (S8). In contrast the confirmation of *D* (as part of *DE*) in S8 is actually negative; thus no measure of confirmation (difference, ratio, etc.) would predict that *D* should be involved in the preferred hypothesis.—Similarly, in *S9* and *S10* the frequency in t3 was identical; however, the selection of hypotheses was in line with priors and BL but not with a confirmation approach. In S9, the mostly selected *B* answer is not confirmed but actually disconfirmed in t3 (relative to t1 and t2) (67% relative to 79%), whereas *A* and *C* were confirmed. In contrast, in S10, with most ADE selections, *B* is actually confirmed and *A* and *C* disconfirmed. Thus confirmation seems unable to account for the data.—*S11* corroborates that the most participants change hypotheses if a new pattern in t3 is clearly predicted (here *E*). However, some selected the

Panel 1: Hypotheses Selected to be Most Probable for Samples, Phase 1

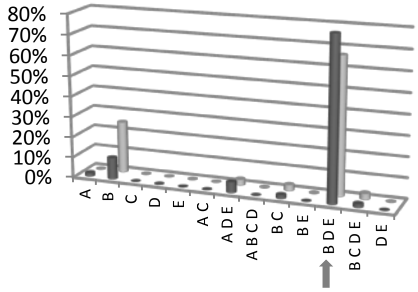
S1

A	B	C	D	E
16	5	3	5	4



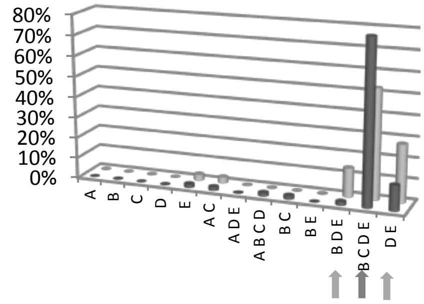
S2

A	B	C	D	E
2	14	1	9	9



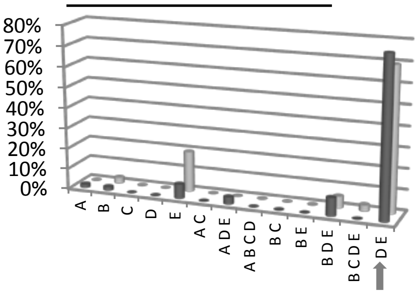
S3

A	B	C	D	E
2	17	14	25	33



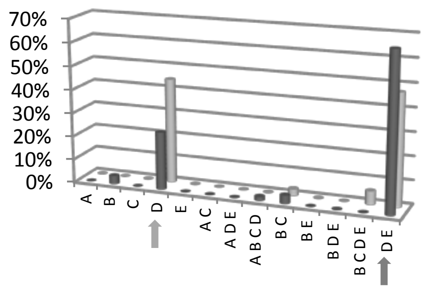
S4

A	B	C	D	E
1	6	0	29	59



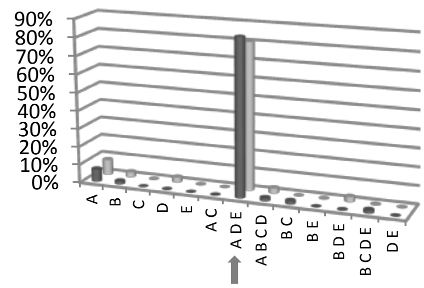
S5

A	B	C	D	E
2	2	1	17	9



S6

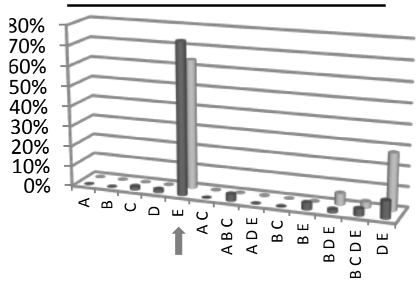
A	B	C	D	E
30	3	4	20	15



Panel 2: Hypotheses Selected to be most Probable after Successive Sampling, Phase 2

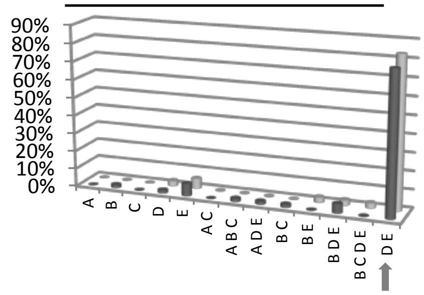
S7

	A	B	C	D	E
t1	4	7	2	6	71
t2	6	7	5	4	79
t3	0	0	0	1	2



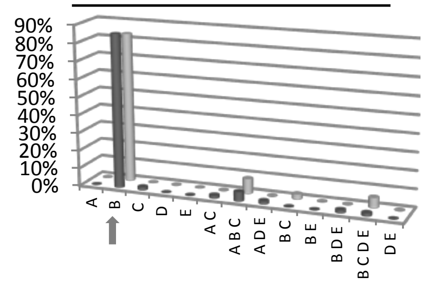
S8

	A	B	C	D	E
t1	0	3	3	33	40
t2	4	6	0	38	44
t3	0	0	0	1	2



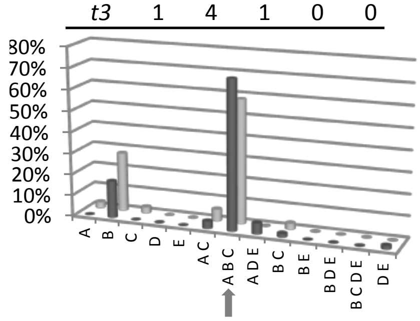
S9

	A	B	C	D	E
t1	2	22	0	2	2
t2	3	31	2	3	0
t3	1	4	1	0	0



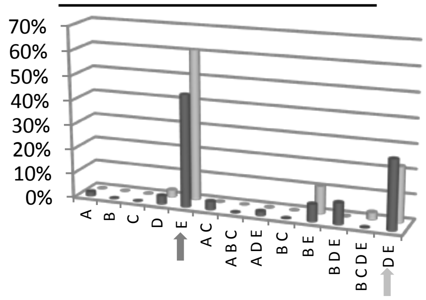
S10

	A	B	C	D	E
t1	13	12	14	0	1
t2	22	13	14	1	2
t3	1	4	1	0	0



S11

	A	B	C	D	E
t1	1	0	3	39	3
t2	2	0	3	33	1
t3	1	3	0	1	34



S12

	A	B	C	D	E
t1	3	20	1	39	30
t2	2	21	3	33	40
t3	1	11	2	9	11

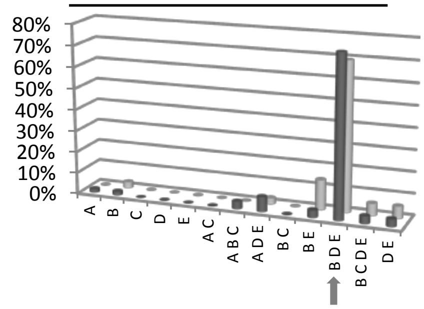


Figure 2: Proportions of selected hypotheses to be most probable in different schools (*S*) in Exp. 1 (jobs: dark) and Exp. 2 (political attitude: grey). Arrows refer to the same main (and further minor) BL-predictions in both experiments.

DE hypothesis, thus putting stronger emphasis on the prior than expected. This is even more at odds with confirmation, since *D* was disconfirmatory.—Finally, *S12* corroborates the predicted hypothesis of BL (*BDE*), whereas classes *D* and *E* were actually disconfirmatory.

General Discussion

Overall, the results obtained in these first frequency-based experiments on monadic polytomous BL strongly corroborate BL's predictions. People commit inclusion fallacies even within one dimension with many subclasses. We found only few or minor differences between (a) label conditions, (b) AND vs. OR formulations, and (c) the experiments with different content. The differences between the two experiments—if not entirely due to chance—may be explained by differences in noise-tolerance, which can be modelled within BL. The results strongly support a pattern-approach of IFs and could not differ more clearly from extensional probability. Here non-*A* involved several subclasses and people judged $P(A) > P(\text{non-}A)$ even if $f(A) > f(\text{non-}A)$ was the case (dissociation with frequencies).

To my knowledge, the present results cannot be explained by any of the many alternative accounts of conjunction fallacy (see von Sydow, 2011; von Sydow & Fiedler, 2013). Tentori et al. (2011) correctly note that most approaches of CFs cannot account for a dissociation with frequencies. Although a confirmation approach may do so, it cannot account for our results: in Phase 1, there was no 'inverse' influence of prior beliefs, as predicted by confirmation. If one alternatively assumes flat priors, since each situation is evaluated anew (as we assume), confirmation in contrast to the results would always predict the selection of very specific hypotheses (with modal frequencies). Moreover, Phase 2 explicitly addressed priors, by varying previous information about the schools. Although people as predicted preferred using newer evidence, they were influenced by the priors if the new evidence did not clearly show that one is concerned with a population that differs from the old one. Crucially, the effect of priors went in the direction predicted by our Bayesian pattern approach, not in the opposed direction of a confirmation approach. Although these results may shed light on previous findings (apparently favouring confirmation; Tentori et al., 2013), a disconfirmation of a confirmation account does not imply its falsification. Confirmation may nonetheless play a role to explain one class of CFs (even if one distinguishes it from pseudo-confirmation effects, such as scale construction effects, or relevance effects). Generally, inclusion fallacies presumably have several different causes, including misunderstanding of terms (Hertwig, Benz, & Kraus, 2008; von Sydow, 2014), unclear sets (Sloman, Over, Slovak, & Stibel, 2003), confusing confirmation with probability (Lagnado & Shanks, 2003; Tentori et al., 2013), or using pattern probabilities where CFs cease to be fallacies (von Sydow, 2001, 2014). *Confirmation* of *patterns* may even play a role.

The present results, in any case, corroborate a pattern-probability account. The findings nonetheless suggest that standard dichotomous BL (von Sydow, 2001) only applies if one is concerned with two classes and dichotomous events. The results suggest that a proper formalization of pattern-probabilities depends on the subjective representation of classes—the intension—, not only on the extension. If one is concerned with stories, as in the original Linda scenario, such representations may vary in an uncontrolled way. Formalizations of BL (and of heuristic approximations of pattern probabilities) should take subjective representations into account (for another aspect, see von Sydow, 2014).

In conclusion, the present research corroborates BL in a new domain. It reveals new avenues of research and suggests that the role of representation in probability judgments is richer than previously assumed. Subsequent research is crucial to address such issues in more detail.

Acknowledgments

This work was supported by the grant Sy111/2-1 from the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program *New Frameworks of Rationality* (SPP 1516). I am grateful to Niels Braus, Caroline Nettekoven, and Maria Waltmann for running the experiments and to Martha Cunningham for corrections. Furthermore I would like to thank anonymous reviewers for very helpful comments and suggestions.

References

Fiedler, Klaus & von Sydow, Momme (2015). Heuristics and Biases: Beyond Tversky and Kahneman's (1974) Judgment under Uncertainty (pp. 146-151). In: M. W. Eysenck & D. Groome. *Cognitive Psychology: Revisiting the Classical Studies*. Los Angeles, London: Sage.

Foley, R. (2009). Beliefs, Degrees of Belief, and the Lockean Thesis. In: F. Huber, C. Schmidt-Petri (eds.), *Degrees of Belief*, Synthese Library 342, Heidelberg: Springer.

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

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

Kahneman, & Tversky (1996). On the Reality of Cognitive Illusions. *Psychological Review*, *103*, 582-591.

Lagnado, D. & Shanks, D. (2003). The influence of hierarchy on probability judgments. *Cognition, 89*, 157–178.

Navarro, D. Dry, M. J. & Lee, M. (2012). Sampling Assumptions in Inductive Generalization. *Cognitive Science, 36(2)*, 187–223.

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

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

Tenenbaum, J. & Griffiths, T. (2001a). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24(4)*, 629-640.

Tentori, K., Crupi, V., & Russo, S. (2013). Determinants of the conjunction fallacy: Confirmation versus probability. *Journal of Experimental Psychology: General*, *142*, 235-255.

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

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

von Sydow, M. & Fiedler, K. (2012). Bayesian Logic and Trial-by-trial Learning. *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society* (pp. 1090 - 1095). Austin TX: Cognitive Science Society.

von Sydow, Momme (2014). Is there a Monadic as well as a Dyadic Bayesian Logic? Two Logics Explaining Conjunction 'Fallacies'. *Proceedings of the Thirty-Sixth Annual Conference of the Cognitive Science Society* (pp. 1712-1717). Austin TX: Cognitive Science Society.