



HAL
open science

Deductive Schemas with Uncertain Premises Using Qualitative Probability Expressions

Guy Politzer, Jean Baratgin

► **To cite this version:**

Guy Politzer, Jean Baratgin. Deductive Schemas with Uncertain Premises Using Qualitative Probability Expressions. *Thinking and Reasoning*, 2016, 22 (1), pp.78-98. 10.1080/13546783.2015.1052561 . ijn_01293776

HAL Id: ijn_01293776

https://hal.science/ijn_01293776v1

Submitted on 1 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

This is an Accepted Manuscript (10 May 2015) of an article published by Taylor & Francis in *Thinking and Reasoning*, available online:

<http://www.tandfonline.com/doi/full/10.1080/13546783.2015.1052561>

**Deductive Schemas with Uncertain Premises
Using Qualitative Probability Expressions**

Guy Politzer^a and Jean Baratgin^{a,b}

^aInstitut Jean Nicod, Ecole normale Supérieure, Paris, France

^bCHArt (PARIS), Université Paris 8, France

Abstract

The new paradigm in the psychology of reasoning redirects the investigation of deduction conceptually and methodologically because the premises and the conclusion of the inferences are assumed to be uncertain. A probabilistic counterpart of the concept of logical validity and a method to assess whether individuals comply with it must be defined. Conceptually, we used de Finetti's *coherence* as a normative framework to assess individuals' performance. Methodologically, we presented inference schemas whose premises had various levels of probability that contained non-numerical expressions (e.g., "the chances are high") and, as a control, sure levels. Depending on the inference schemas, from 60% to 80% of the participants produced coherent conclusions when the premises were uncertain. The data also show that (i) except for schemas involving conjunction, performance was consistently lower with certain than uncertain premises, (ii) the rate of conjunction fallacy was consistently low (not exceeding 20%, even with sure premises), and (iii) participants' interpretation of the conditional agreed with de Finetti's "conditional event" but not with the material conditional.

Key words: uncertain reasoning; deduction; de Finetti's coherence; conjunction fallacy

Traditionally, to investigate human deduction psychologists relate human rationality to binary classical logic. However, some authors (Oaksford & Chater, 2007, 2009; Over, 2009) have argued that logic is inadequate to account for performance in reasoning tasks because people use strategies to reason under uncertainty whose nature is probabilistic. These authors propose to change the reference of the normative system by abandoning reference to binary logic altogether and adopting Bayesian probability. In this *new paradigm* (Elqayam & Over, 2013; Evans, 2012; Evans & Over, 2013; Over, 2009) the normative model of binary logic is replaced with a Bayesian model (providing a "probabilistic logic") making it possible to study the uncertain deductive inferences which people carry out in daily life. Of course, there are various concepts of Bayesianism. The literature in Philosophy, Logic, Linguistics, Economics, and Artificial Intelligence offers a wide range of options and the choice that is made for normative or descriptive purposes has important consequences (Baratgin, 2002; Baratgin & Politzer, 2006, 2007; Elqayam & Evans, 2013; Mandel, 2014). The present study makes a radical choice by adopting subjective Bayesian probability (de Finetti, 1964). The theoretical and methodological relevance of de Finetti's Bayesian theory as a model of rationality is outlined in Baratgin & Politzer (in press). Theoretically, this approach hinges upon the notion of Bayesian coherence which stipulates that degrees of belief must respect the axioms of probability (static coherence) and that the revision of degrees of belief must follow the conditioning principle (dynamic coherence).¹

Bayesian coherence allows to define human rationality in a context of vague or partial knowledge, unifying in the same framework deduction and probability judgment. This theoretical advantage carries over to the methodology, which can be applied indifferently to these two domains.

De Finetti's approach distinguishes two levels of experimental analysis corresponding to the two subjective levels of knowledge of an event (de Finetti, 1980). We define these below.

(i) An *elementary* level of knowledge which is the level of the individual's belief about the realization of the event. A specific object or phenomenon is defined by the characteristics of

the phenomenon that are known to the individual. An event C is always conditioned on the individual's personal state of knowledge, that is, an event can always be described as a *conditional event* denoted $C|A$. For de Finetti the value *true* attributed to an event corresponds to its realization and the value *false* to the absence of its realization. There is a third value U that corresponds to the uncertainty about the realization of this event. Recent experimental results show that indeed a majority of individuals have a trivalent interpretation of the conditional (Politzer, Over & Baratgin, 2010) and that the logic of reference is a tri-valued logic consistent with de Finetti's (1995) logic, uncertainty being represented by an additional truth value (Baratgin, Over & Politzer, 2013; 2014).

(ii) A *meta-epistemic* level of knowledge that concerns *degrees* of belief about the event C .

It involves the individual's subjective additive probabilities. At this level the value U just mentioned may be specified in probabilistic terms. For instance in Jeffrey's (1991) approach it equals $P(C|A)$.

Most experimental studies have focused on the probability of indicative conditionals *if A then C* in natural language, and found that, as expected, it is interpreted as a conditional probability (Cruz & Oberauer, 2014; Evans & Over, 2004; Oaksford & Chater, 2007, 2009; Pfeifer & Kleiter, 2010; Politzer, Over & Baratgin, 2010). However, relatively few studies have investigated human coherence in deduction under uncertainty in which the event under consideration is the conclusion of an argument (Cruz, Baratgin, Oaksford & Over, 2015; Evans, Thompson & Over, 2015; Pfeifer, 2014; Pfeifer & Kleiter, 2009, 2010, 2011; Singmann, Klauer, & Over, 2014).

De Finetti's Coherence

The main goal of this paper is to study people's performance on a few elementary inferences for which data are scarce, in line with the methodology used by de Finetti (1964) to appraise the coherence of probability assessments (see also Hailperin, 1996, 2010). It consists of comparing participants' assessments of the conclusion with the coherence intervals determined by the probability of the premises. This methodology has the advantage of offering a unified framework to study both deduction under uncertainty and probability

judgment (Baratgin, in press; Baratgin & Politzer, in press).²

De Finetti's (1964) *coherence* approach to subjective probability refers to the betting interpretation of probability assessment. A probability assessment is coherent if it does not admit one or more bets with sure loss. De Finetti's (1974, 1975) fundamental theorem of probability states that given the probabilities of a set of events, one can derive the coherent probability of a further event using the probability calculus. This applies in particular to a set of events expressed by the premises of an argument and to the event expressed by the conclusion. There are two possible results. One is the whole $[0, 1]$ interval, in which case the argument is called "probabilistically non informative" (Pfeifer & Kleiter, 2006a). The other is a narrower interval $[l, u]$ in which case the argument is called "probabilistically informative"; the closer the limits are to each other, the more precise the conclusion; as a special case the limits may be equal, in which case the conclusion is a precise point probability.

This is the approach advocated by Coletti and Scozzafava (2002) and Pfeifer and Kleiter. These authors (Pfeifer & Kleiter 2009, 2010, 2011) have developed a research program that aims to test the hypothesis of a "mental probability logic" as a model of performance. They studied a number of inference schemas in which the premises had the form "at least P% and at most Q% of [set] have/are [property], for instance, referring to students, "at least 25%, and at most 55% have black hair".

For the rule of *and-introduction* (from A; C, infer A and C) a second premise referred to another property with different percentages and participants were asked to determine the percentage who had both properties; they could answer by giving either a point percentage or the limits of an interval. Two experiments (Pfeifer & Kleiter, 2005) using different sets and properties gave similar results: 64% and 76% of the answers were within the coherence interval.

The same methodology (Pfeifer & Kleiter, 2007) was applied to study *Modus Ponens* and *Denying of the Antecedent*. The premises were of the type "exactly 80% of the red cars on this parking lot are two-door cars"; the minor premise for MP was, for instance: "exactly 90% of the cars on this parking lot are two-door cars", and the question: "Imagine all the cars that

are on this parking lot. How many of these cars are two-door cars?" Sixty-three percent of the participants gave coherent intervals for MP but only 41% for DA, (a difference which the authors explained by matching-based guessing with DA). The result for MP is comparable with that of the pioneering study by George (1997, experiment 2) who presented the minor or the major or both premises of MP arguments as "very probable" and observed 70% of responses which, with hindsight, can be identified as coherent (see also Evans & al., 2015, and Singmann & al., 2014 for similar results).

Pfeifer & Kleiter (2006) studied also *contraposition* with a premise of the type "exactly 93% of the cars on a parking lot are blue", and the question, "Imagine all the cars that are not blue. How many of the non-blue cars are not on the parking lot?" Pooled over two experiments about 50% of the participants gave wide intervals suggesting that they appreciated that the inference is non-informative. However, for the rule of *transitivity* (*if A then B; if B then C; therefore if A then C*) a majority responded by point values indicating that they incorrectly endorsed the rule as informative, which the authors attribute to the misinterpretation of the rule as the *CUT* rule (*if A then C; if A and B then C; therefore if A then C*) due to a conversational implicature.

We have two comments to make on these results. One, for the informative schemas, there exists a clear majority of individuals who give coherent probability intervals but this is not a large majority. The hypothesis that individuals can appreciate the coherence intervals of the conclusion of simple deductive schemas is only moderately supported, unless one can invoke performance factors (linguistic, or linked to the interpretation of the task, etc.) which add noise to the data, a problem that is pervasive in reasoning research. This is a sufficient reason to extend research and gather more data.

Two, there are at least two features of the methodology that could influence the results. The first one is that the conditional sentences which constitute most of the schemas investigated are not expressed with an *if-then* construction but instead with "N% of the X are Y". This may force an interpretation of the sentence, and whether it is appropriate to communicate de Finetti's conditional event is an open question. The other feature is the use

of numerical values to communicate probabilities. Their usage (be they values expressed between 0 and 1 or as percentages) is the result of scientific development and educational acquisition at the social and individual levels, respectively. At the root of the psychological investigation of probabilistic reasoning there is the assumption that individuals have a sense of the probable, and at the root of the new paradigm there is the assumption that this is a basic cognitive ability. The fact that degrees of belief are indeed verbalisable in language by many expressions suggests that these constitute a natural and appropriate format to communicate probability. We view the fact that they are imprecise as a reflect of the manner in which they could be mentally represented. From this point of view, the precision of numerical estimates is illusory: if one wonders whether the Brazilian soccer team will pick up for the next world cup, it makes little sense to discuss whether the chances are, say, 62% rather than 65%.

The study that will be reported aims to examine individuals' coherence in their evaluation of the conclusion of a few deductive schemas using non-numerical probability expressions and the *if-then* formulation of conditional sentences. The schemas are probabilistically informative with one exception. They were chosen because they are basic in systems of natural logic and they allow the determination of the coherence intervals using non numerical expressions. They are listed below together with their coherence intervals. A general method to determine the coherence intervals (by solving linear equations) can be found in Coletti and Scozzafava (2002) and in Gilio (2002). For the elementary schemas that are considered here an easier method can be found in Politzer (in press).

AND-elimination: $A \text{ AND } C \quad \therefore \quad A$

$$P(A) \in [P(A \text{ AND } C), 1]$$

AND-introduction: $A; C \quad \therefore \quad A \text{ AND } C$

$$P(A \text{ AND } C) \in [\max \{0, P(A)+P(C)-1\}, \min \{P(A), P(C)\}]$$

$$P(A \text{ AND } C) \in [\max \{0, 2b-1\}, b] \text{ when } P(A) = P(C) = b$$

OR-introduction: $A \quad \therefore \quad A \text{ OR } C$

$$P(A \text{ OR } C) \in [P(A), 1]$$

AND to IF: $A \text{ AND } C \therefore \text{IF } A \text{ THEN } C$

$$P(\text{IF } A \text{ THEN } C) \in [P(A \text{ AND } C), 1]$$

OR to IF-NOT: $A \text{ OR } C \therefore \text{IF NOT-}A \text{ THEN } C$

$$P(\text{IF NOT-}A \text{ THEN } C) \in [0, P(A \text{ OR } C)]$$

contraposition: $\text{IF } A \text{ THEN } C \therefore \text{IF NOT-}C \text{ THEN NOT-}A$

$$P(\text{NOT-}A|\text{NOT-}C) \in [0, 1]$$

The determination of the bounds using non-numerical expressions is straightforward. The bounds either reproduce the probability of the premise or are equal to 0 or 1. There is one exception with *AND-introduction* for which the lower bound involves an addition. For the sake of simplicity it was decided to give the same probability b to the two premises, so that the upper bound $\min\{a, c\}$ is reduced to the common value b , and the lower bound $\max\{0, a+c-1\}$ can be defined as follows. When $a = c$ is lower than *average* (*low, very low*) the bound is 0; when $a = c$ is *average* the bound is 0 or *very low*; when $a = c$ is *high* the bound is *average*; and when $a = c$ is *very high* the bound is *high* or *very high*.

Experiment

Method

Material

Constitution of the items. The six inference schemas were formulated as follows: "Knowing that the chances that [premise(s)] are [level of probability], "in your opinion, what are the chances that [conclusion]"? There were seven levels of probability: five uncertain formulated as *very low, low, average, high, very high*, and two non-probabilistic with chance levels of 0% and 100% (that will be called "sure" conventionally). This was followed by a multiple-choice response format. For the sure cases there were two options: "exactly 0%"; "greater than 0%" in the former case, and "exactly 100%"; "smaller than 100%" in the latter case. For the uncertain cases, three options (henceforth called *primitive* options) were offered: "greater than [premise level]", "just [premise level]", and "smaller than [premise level]"

with detailed instructions explaining that any number of options (one, two, or all three) could be selected. In this way the following seven possible responses could be obtained after possible combination: *greater; equal; smaller* (only one primitive option chosen); *greater or equal; smaller or equal; greater or smaller* (two primitive options); and *greater, equal or smaller* (all three primitive options); these will be called *derived* options.

Content. Two kinds of content were used. The *chips* content referred to chips that could be drawn randomly. Here is an instance of item (translated from French) for AND-elimination with a 0% level of belief in the premise:

Pierre and Marie are playing with a bag of chips. They are drawing chips at random. Knowing that the chances for *Pierre to draw a chip that is square and white* are 0%, in your opinion, the chances for *Pierre to draw a square chip* are: greater than 0%; exactly 0%.

The *activity* content referred to the whereabouts of two persons. Here is an instance of item for the OR to IF-NOT inference with a high level of belief in the premise:

Knowing that the chances are high that now *Nicolas is in Lyon or Jeanne is in Marseille (or both)*, in your opinion, the chances that now *if Nicolas is not in Lyon, Jeanne is in Marseille* are: greater than high; just high; smaller than high.

Procedure

Mode of presentation of the material and design. The items were presented in booklets. The first page of the booklets contained the instructions which were followed by three schemas (one per page). The choice of the schemas was constrained in such a way that AND-elimination and AND-introduction did not appear in the same booklet; also the same connective did not appear in more than two schemas in a same booklet; this constraint on the selection of the schemas resulted in the partition of the schemas into two groups and corresponding constitution of two types of booklets. Orthogonally, the choice of the levels of belief yielded seven types of booklets as follows. For the two sure levels the first and third schemas were at one level and the second schema was at the other level (two types). For

the five uncertain levels, booklets with one *low* between two *high* levels, and vice versa were constituted (two types); similarly for the *very low* and *very high* levels (two types); and booklets with a sequence of three *average* levels (one type). Finally, the two kinds of content constituted a between-subjects variable. All this resulted in the constitution of 28 types of booklets (2 groups of schemas x 7 combinations of levels of probability x 2 kinds of content).

Participants. A total of 754 students from the Universities of Paris 6, Paris 8, Paris 13 and the EPITA school (mean age = 25.8) participated in the experiment on a voluntary basis. The great majority were students of social sciences, and none of them had significant knowledge of logic or probability theory. They were all native speakers of French. The administration took place in small groups of three or four. For a given schema at a given level of belief and a content type, the number of observations varied from 18 to 31 with a mean of 26.

Results

Defining the coherent responses. A response is deemed coherent if and only if it belongs to the coherence interval of the schema under consideration. For instance, if the coherence interval spans from *equal* to *greater*, then any one of the three responses *equal*, *greater*, and *equal or greater* is coherent. Note that there is a subtle difference between the two responses *equal* and *greater* on the one hand, and the response *equal or greater* on the other hand: each of the former constitutes a part of the interval whereas the latter covers the whole of the interval. At first sight, it might be thought that the only correct response is the latter. However, from a subjectivist point of view (de Finetti, 1964), many of the individuals' beliefs to which the theorist has no access contribute to the evaluation of the conclusion over and above the degree of belief in the premises, so that the individuals' evaluations are constrained and have no reason to cover the whole interval. In brief, individuals are rational provided their evaluations fall inside the coherence interval, which is tantamount to respecting the laws of the probability calculus. In case the whole interval is chosen, this may result also from the current state of beliefs of the individual or from a formal approach to the task reflecting metacognitive awareness.

Presentation of the results. Table 1 gives a summary of the results. It presents, for each of the six schemas, the distribution of the coherent responses as a function of the premise probability. The figures are given in percent rounded to the closest unit, collapsed across content (because the two values were generally close, see below) and across responses (such as *greater*, *equal*, etc.) The detailed results showing the cross distributions of participants' choices over the probability of the premises and the levels of belief in the conclusion are given in Table 2 in the Appendix. Due to the difference in response formats, the uncertain condition and the sure condition will be analysed separately.

Table 1. Distribution of the coherent responses (in percent) as a function of the probability of the premise for each of the six schemas.

Inference schemas	P r e m i s e c h a n c e s					Mean of columns 1 to 5	0%	100%
	very low	low	average	high	very high			
AND-elimination	71	91	96	79	84	84	100	81
AND-introduction	90	88	91	80	60	83	81	94
OR-introduction	73	80	84	81	85	81	100	58
AND to IF	67	78	74	90	86	79	100	69
OR to IF-NOT	69	78	78	83	79	78	54	100
Contraposition	100	100	100	100	100	100	100	100

The uncertain condition. Five schemas have three levels of evaluation that are coherent (the top three levels for *AND-elimination*, *OR-introduction* and *AND to IF*) and the bottom three levels for *AND-introduction* and *OR to IF-NOT*) while all the levels are coherent for contraposition. The derived option *greater or smaller* was never chosen. A first assessment

for the five probabilistically informative schemas can be made: the mean percentage of coherent evaluations collapsed across levels of belief in the premises and across schemas indicates a rate of coherence of 81%, with little variability across schemas and material. Considering the five levels of premise probability crossed with the five schemas, which yields 25 distributions, the rates of coherent responding ranged from 65% to 96%. However suggestive these figures are, we need to examine the data more precisely after defining an appropriate random model of responding against which the coherence model will be compared.

Keeping aside contraposition for the time being, under the coherence hypothesis the schemas are expected to lead to a response pattern that consists of three options (the derived options) out of seven, viz. *greater; greater or equal; equal* for *AND-elimination*, *AND to IF* and *OR-introduction*; and *smaller; smaller or equal; equal* for *AND-introduction* and *OR to IF-NOT*. A first coarse choice for the null hypothesis consists of assuming that participants do not attempt to answer the question and choose among the options randomly. Considering the seven possible derived options as equiprobable, the probability for an individual to select by chance a derived option that belongs to the coherent interval is $3/7$. But this model does not seem satisfying because it is unlikely that participants select one, two or three primitive options with the same probability. It seems more realistic to attribute weights w_i such as 1 for the 3-option response, 2 for a 2-option response, and 3 for a 1-option response (meaning, for instance, that the ratio of the probability of selecting all three options to selecting one single option is 1 to 3, etc.) for a total weight $\sum w_i n_i = (1 \times 1) + (2 \times 3) + (3 \times 3) = 16$. According to this pseudo-random model, because the coherent response patterns always consist of one 2-option response and two 1-option responses, the probability for such a pattern to be chosen is $(1 \times (2/16)) + (2 \times (3/16)) = 1/2$.

However this model, like the equiprobable random model, is still unrealistic because it is based only on the number of derived options and their combinatorial complexity, and disregards the meaning of the options. We can make a further improvement by considering that participants answer also on the basis of the denomination of the primitive options (while

still ignoring the content of the item as previously). In that case the *smaller or greater* response is unlikely to be chosen because it may seem internally contradictory. We modify the weights accordingly by attributing also a weight of 1 to the latter response, hence a total weight of 15 and a value of $(1 \times (2/15)) + (2 \times (3/15)) = 8/15 = 53.3\%$ as a percentage of reference. Notice that in choosing this value rather than the two preceding ones (43.3% and 50%) we take a conservative stand and decrease the risk of type 1 error.

The mean percentages of coherent evaluations of the conclusion (after collapsing across the levels of probability of the premise) for the five inference schemas and both material types are indicated in each subtable of Table 2; they range from 74% to 89%. They all differ from the 53.3% value of reference (all chi-squares with $df = 1$ are highly significant).

Considering now each level of premise probability separately, we have $5(\text{schemas}) \times 2(\text{materials}) \times 5(\text{levels}) = 50$ comparisons to test (the N are indicated in the subtables in the Appendix; they range from 18 to 31 and have a mean of 26). The rate of coherent evaluations range from 54% to 96% (with an average of 81% as mentioned above). Because each participant responded to several schemas, the percentages were compared to the baseline using z tests corrected by the Bonferroni-Holm method for family-wise type 1 errors. They differ from the 53.3% value at the level of significance of $p < .05$ in 41 cases; in the remaining nine cases the difference is always in the expected direction. The nine non-significant cases concern each of the five schemas at least once and at most twice, that is, the schemas are roughly equally concerned; six of these nine cases concern the *very low* level of belief in the premise. We now review the schemas in turn. Because the difference in the *mean* rates of coherence between the two materials did not exceed 10% on any of the five cells, we consider the mean of the two relevant percentages.

AND-elimination. About 84% of the responses were coherent: most participants appreciated that a conjunct could have a probability greater, greater or equal, or equal to the conjunction and very few estimated that it could have a probability smaller or equal, smaller, or indeterminate.

AND-introduction concurs with the preceding schema as about 83% of the responses were coherent and symmetrically most participants appreciated that the conjunction could have a probability smaller, smaller or equal, or equal to a conjunct while the remaining estimated that it could have a probability greater, equal, or indeterminate. In brief, the results of the two *AND* inferences are very consistent and concur to demonstrate that the rate of fallacy is limited to about 20%.³

OR-introduction. The 81% rate of coherent responses shows that most participants were aware that the probability of a disjunction can only be equal or greater than that of a disjunct. This is in sharp contrast with the result of the classic sure case (confirmed here with the 100% level for a strong minority, see below) that individuals are reluctant to infer A OR C from A.

AND to IF. Most participants (79%) were coherent and correctly had equal or stronger belief in IF A THEN C than in A AND C.

OR to IF-NOT. Most participants (78%) were coherent and correctly had equal or weaker belief in IF NOT-A THEN C than in A OR C. The modal response was *equal*, for each level of probability and for both materials; this suggests that a relative majority of participants may have been insensitive to the decrease in probability. Following Gilio & Over (2012) this is what can be expected when the premise *A or C* is interpreted in a non-constructive way, which is likely to be the case in our impoverished scenarios that do not offer any reason to strongly believe any one of the disjuncts separately.

We now turn to the case of *contraposition* for which the coherence interval contains all the derived options, hence a predicted choice of all these options, which is also the prediction based on random responding. However, it is worth examining whether, and in the affirmative how, the observed distribution over the seven derived options differs from the pseudo-random distribution. It does differ for both the *chips* and the *activity* materials (chi-square = 43.4 and 65.5, $p < .01$, $df = 6$). A close examination of the distributions is suggestive (Table 2f). One, for both materials, the relatively high percentages in rows 3 to 5 suggest an *equal to smaller* interval. To test this model we sum the three relevant

percentages which add up to 62% and 78% respectively; they differ significantly from the 53.3% value (chi-square = 4.1, $p < .05$ and 31.1, $p < .001$, $df = 1$). Two, it is remarkable that for the chips material the last row (*smaller, equal or greater* option, that is, the choice of the whole interval), contains an average of 22% of choices, which differs significantly from 1/15 (chi-square = 50.1, $df = 1$, $p < .0001$); this result is to be compared with the other five inferences whose percentages on the last row range from 1% to 9%.

We interpret these results as follows. It seems that for both materials a majority of participants tended to treat contraposition not as a probabilistically indeterminate inference but as an inference whose probability of the conclusion could be evaluated as lower or equal to that of the premise (like *AND-introduction* and *OR to IF-NOT*). However, a minority of participants with the chips material were aware that the coherence interval is the whole interval. There is one explanation for the bias toward lowering the probability of the conclusion, admittedly *post hoc*, but testable. When the additional premise A is introduced, it can be shown (Politzer, 2014) that under the assumption that $p(A)$ is high (typically $p(A) > .75$), and for values of the conditional not too low (typically $P(C|A) > .25$), the probability of the contrapositive is lower than the probability of the conditional. So, the observed bias could coincide with the implicit assumption that $p(A)$ is high (meaning that a number of participants did not solve precisely the inference presented to them but assumed the additional premise A).

Another possibility to explain why participants had a tendency to decrease the probability of the conclusion is that, while being aware of the non-validity of the inference, some people found it more informative to designate the lower levels of chance (which are diagnostic of non-coherence) rather than all the levels which include higher coherent levels.⁴

So far, we have pooled the percentages across the levels of probability of the premises. We did not make any predictions regarding a possible effect of this variable on the coherence of the evaluation of the conclusion. Inspection of the last row of the tables indicates that, with one exception, there are only unsystematic increases and decreases in the rate of coherent responses but no systematic trend as a function of the probability of the

premise. The exception concerns *AND-introduction* where there was a trend for the *high* and *very high* levels of probability to yield a rate of coherent responses inferior to the other levels. The rate of fallacy for the two higher levels (69%) was significantly lower than for the other three levels pooled together (89.3%) (chi-square = 16.8, $p < .001$, $df = 1$, $N = 271$). We have no explanation for this result.

The sure condition. For each of the six schemas there were two primitive options (*equal* and *greater* for the 0% level, *equal* and *smaller* for the 100% level) and therefore three derived options. The deductively valid conclusions are, depending on the schema and the level, either the *equal* option or all three options. They appear shaded in Tables 2a through 2f. To evaluate the responses we need again to define a realistic model of random responding. Giving the 1-option responses (*smaller*, *equal*, *greater*) a weight that is twice as great as the weight of the 2-option response (*greater or equal*, *smaller or equal*) the percentage of reference for the valid *equal* response is $(2/(2+2+1)) = 40\%$; in other words, we take this value as a norm to evaluate the significance (at the level of $p < .05$) of the observed rates of the 1-option responses. Note that in choosing this weight (rather than equal weights leading to one-third for each of the three derived options), we take again a conservative stand. The results are strikingly similar for the two types of material and they will not be distinguished, unless there is an exception. All the significant differences refer to z tests at a level of $p < .05$ after using the Bonferroni-Holm correction whenever applicable.

For *AND-elimination* with chances of 0% a majority chose the *greater* response, showing that many overlooked the possibility that the chances could be null for the other conjunct C; with chances of 100% a large majority chose the correct *equal* response.

For *AND-introduction* an overwhelming majority chose the correct *equal* response with chances of 100% (and a strong majority chose the correct *equal* response when the chances are 0%). This result is remarkable because it is a measure of the conjunction fallacy with sure premises, and its order of magnitude is the same as the mean rate observed with uncertain premises, namely about 20%. This relatively low rate in an inferential task is at variance with the classic data obtained in a non inferential framework.

For *OR-introduction*, with chances of 0% a majority chose the *greater* response, showing that many overlooked the possibility that the other disjunct (C) could be null. With chances of 100%, with the *activity* material the majority chose the correct response but with the chips material there was no difference with random responding; this apparently confirms the well documented difficulty of the non probabilistic version of this inference, which will be discussed below.

For *AND to IF*, with chances of 0% all three responses were predicted; in fact there is a significant preference for the *equal* response; with chances of 100% a majority chose the predicted *equal* response with the chips material but this failed to reach significance with the activity material.

For *OR to IF-NOT*, with chances of 0% the distribution did not differ from chance responding; with chances of 100% the pattern of responses contained all three options, which is correct according to the conditional event construal of the conditional sentence, but does not differ from random responding, so that this result is inconclusive.

For *contraposition* participants' responses were as predicted spread over all the options. When the premise probability was 0% there was a preference for the *greater* response; when the premise probability was 100% a minority chose the *equal* (100%) response while the majority chose the *smaller* response (as they did in the uncertain case). It is noticeable that with our material in the classic situation where a conditional is presented as certain only a minority inferred the contrapositive.

Leaving aside contraposition (for which we have noticed an effect of the material) the 5 (schemas) x 7 (premise probability) x 6 (response options) define 210 cells in the tables, out of which only 12 show a difference in the rates of response between materials significant at the .05 level; this number is close to what can be expected by chance (210 x 5%).

Furthermore, after collapsing across premise probability (and so considering only the 30 resulting cells), there is not a single case of significant difference between materials. We cannot conclude to any systematic effect of the materials and this speaks in favour of the robustness of the results.

General discussion

We begin with the main question that motivated this study, and then review some auxiliary results concerning conjunction, the sure situation, and the interpretation of the conditional.

Coherence

To begin with, the observation that across five schemas about 80% of participants gave coherent responses leaves open the question of estimating the proportion of participants who gave the correct response by chance, even though all the percentages are statistically significant. In the most favourable situation all the coherent responses would stem from genuinely coherent individuals while the remaining responses would reflect errors committed by the others. In the least favourable situation, the coherent responses would stem from the superposition of a group who follow the pseudo-random responding model and a group of genuinely coherent participants. We can estimate the contribution of the latter as follows. Let γ stand for the observed proportion of coherent responses and β for the proportion who are genuinely coherent; then using the value of $8/15$ defined above for chance responding, the contribution of those who respond by chance is $(8/15)(1 - \beta)$ hence $\gamma = \beta + (8/15)(1 - \beta)$. It follows that $\beta \geq (15/7)(1 - \gamma)$, that is, this is the lowest possible value compatible with the pseudo-random model and the data. For the average value of 80% coherent, this gives for β a minimum of 57%, hence a percentage of correct (coherent) response lying in the range [57%, 80%]. Taking the mid-point of this interval, it appears that about two thirds of our population are genuinely coherent in responding to the five basic schemas under investigation. This is confirmed by considering the schemas and the two kinds of materials separately: the mid-points of the relevant intervals indicate that the coherent rates of response vary between 58% (for *OR to IF-NOT, chips*) and 78% (for *AND-elimination, activity*).

By and large, the rate of coherent evaluations that we observed is equivalent to Pfeifer and Kleiter's observations and this applies to the four schemas not studied hitherto as well as to the *AND-elimination* schema. This answers and dispels the methodological reservations that have partly motivated the present study.

But our main concern remains: Performance that varies from fair to good depending on the inference schemas, with an average of two thirds, is far from perfect. It is confirmed that we are in a state of affairs, common in reasoning studies, in which a sizeable minority of observations do not conform to the predictions of a normative model while the majority do. Independently of specific differences in difficulty between schemas, the errors could be due to individual differences (some part of the population having lower cognitive ability) or to difficulties which affect all participants. Further research is needed to address these questions. We now turn to a few other results that are noteworthy.

Other results

Conjunction

The results for conjunction are very consistent as they show for both *AND* inferences that a stable proportion of individuals committed the conjunction fallacy. However, contrary to most observations reported in the literature based on the standard task such as Tversky and Kahneman's (1983) classic problems, the rate of fallacy was limited to about 20%. Our results concur with Pfeifer and Kleiter's (2005) whose participants were about 70% coherent on *AND-introduction* but produced a rate of 82% of conjunction fallacy on the Linda problem. What can be the origin of this discrepancy? It was not the aim of the present investigation to address specifically the conjunction fallacy, and the design of the experiment allows only some speculation. One hypothesis is suggested by the explanation of the fallacy proposed by Tentori, Crupi, and Russo (2013). They claim that the fallacy depends on the added conjunct (h_2) being perceived as inductively confirmed by the connection with the other conjunct (h_1). In the absence of a scenario, as is the case here, this requires $P(h_2|h_1) > P(h_2)$, which implies probabilistic dependence. But for both materials, there is no reason to assume probabilistic dependence between the two conjuncts (such as *Nicolas is in Lyon* and *Jeanne is in Marseille*) so that the explanation does not apply.

Another hypothesis is that the process of comparison between the probability of the premises and the probability of a conclusion, which results from an inference, may have fostered an attitude of epistemic vigilance (Mercier & Sperber, 2011). Because the essence

of an argument is to provide reasons to believe the conclusion, participants may have been prompted to wonder why the given probability of the premises is a good reason for the conclusion to have or not to have the same probability; and this differs from the simple comparison of the probability of two sentences. Whatever the explanation, our results illustrate that for the fallacy to occur special conditions are required and that the standard task of the conjunction fallacy lacks validity to test individuals' mastery of the probabilistic principle under consideration.

The sure situation

Regarding the sure situation, performance was generally lower than in the unsure situation, except for the two AND inferences. For the *OR-introduction*, studies in the old paradigm showed that the rate of acceptance of the conclusion lies from about 50% to 60% (Braine, Reiser & Romain, 1984; Rips, 1983, Rips & Conrad, 1983). As hypothesised by these authors, this moderate rate can be attributed to a difficulty to integrate the irrelevant information carried by the other, redundant new disjunct C whose introduction contravenes the Gricean cooperative principle or, in other terms, throws out semantic information (Johnson-Laird & Byrne, 1991). This pragmatic explanation seems buttressed by our data for the uncertain case where relatively few participants decreased their belief in the conclusion, which is understandable given the uncertainty of the premise A, for now C is relevant as a possible substitute for A. A similar explanation obtains for the 0% level where indeed the majority gave the response *greater*. In sum, our results suggest that *OR-introduction* belongs to the repertoire of lay reasoners; it goes smoothly under uncertainty but there is a pragmatic obstacle in the case of a sure *forward* inference. (Concerning the sure *backward* inference, which was not studied here, everyday observation indicates that when arguing about the truth of *A or C*, it suffices for reasoners to bring evidence that A to convince their interlocutor that *A or C*, not surprisingly since C is already in the context).

The conditional

Lastly, in supporting de Finetti's theory, the results support the conditional event interpretation of the conditional. It is interesting to examine how the material conditional fares

in comparison. The predictions and the results for *AND to IF* are straightforward: the probability of the material conditional equals $p(A \& C) + p(\text{non-}A)$ (Adams, 1998), which shows that it is greater or equal to the probability of the premise $p(A \& C)$. (The equality occurs when $p(A) = 1$). Therefore the expected responses are the same as the coherent responses. Now for *OR to IF-NOT* and *contraposition* the responses do not always coincide with the coherent responses because both inferences are reversible under the material conditional, so that the probability of the premise and the conclusion are equal, hence an expected response *equal* in both cases for all the levels of probability. For *OR to IF-NOT* the rate of *equal* responses reached only about one-half of the responses in the unsure condition and 43% in the sure (100%) condition. For *contraposition* the results are even worse: the *equal* responses constituted barely more than one third of the responses in the unsure condition and less than one-quarter in the sure (0%) condition. We conclude that the material conditional construal is not supported by the present data. Assuming that for the basic conditionals considered here reasoners build a fully explicit model that coincides with the truth conditions of the material conditional, the result runs counter to the model theory,

References

- Adams, E. (1998). *A primer of probability logic*. Stanford: CSLI Publications.
- Baratgin, J. (2002). Is the human mind definitely not Bayesian? A review of the various arguments. *Cahiers de Psychologie Cognitive / Current Psychology of Cognition*, 21, 653-680.
- Baratgin, J. (in press). *Le raisonnement humain : une approche finettienne* [Human reasoning: A Finettian approach]. Paris: Hermann.
- Baratgin, J. & Politzer, G. (2006). Is the mind Bayesian? The case for agnosticism. *Mind and Society*, 5, 1-38.
- Baratgin, J., & Politzer, G. (2007) The psychology of dynamic probability judgment : Order effect, normative theory and experimental methodology. *Mind and Society*, 5, 53-66.
- Baratgin, J., & Politzer, G. (in press). Logic, probability, and inference: A methodology for a new paradigm. In L. Macchi, M. Bagassi, & R. Viale (Eds.), *Human rationality: Thinking thanks to constraints*. Cambridge, Ma: MIT Press.
- Baratgin, J., Over, D. P., & Politzer, G. (2013). Uncertainty and the de Finetti tables. *Thinking and Reasoning*, 19(3-4), 308-328.
- Baratgin, J., Over, D. P., & Politzer, G. (2014). New psychological paradigm for conditionals and general de Finetti tables. *Mind and Language*, 29(1), 73-84.
- Braine, M. D. S. , Reiser, B. J. , & Rumin, B. (1984). Some empirical justification for a theory of natural propositional logic. In G. H. Bower (Ed.), *The psychology of learning and motivation*. (pp. 313-371). Vol. 18. N. Y. : Academic Press.
- Coletti, G. & Scozzafava, G. (2002). *Probabilistic logic in a coherent setting*. Dordrecht: Kluwer.
- Cruz, N. , Baratgin, J. , Oaksford, M., & Over, D.E. (2015). Bayesian reasoning with ifs and ands and ors. *Frontiers in Psychology*, 6, 192.
- Cruz, N. & Oberauer, K. (2014). Comparing the meanings of “if” and “ all” . *Memory and Cognition*, 42(8), 1345-1356.

- de Finetti, B. (1964). Foresight: Its logical laws, its subjective sources. In H. E. Kyburg Jr. , & H. E Smokler (Eds.). *Studies in subjective probability* (pp. 55-118). New York: John Wiley. [Originally published 1937].
- de Finetti, B. (1974). *Theory of probability*. Vol 1. Chichester: John Wiley.
- de Finetti, B. (1975). *Theory of probability*. Vol. 2. Chichester: John Wiley.
- de Finetti, B. (1980). *Voce Probabilità*. Encyclopedia, 1146-1187. Torino: Einaudi.
- de Finetti, B. (1995). The logic of probability. *Philosophical Studies*, 77, 181-190. [Originally published 1936].
- Elqayam, S ., & Evans, J. St. B. T. (2013). Rationality in the new paradigm: Strict versus soft Bayesian approaches. *Thinking and Reasoning*, 19(3), 453-470.
- Elqayam, S ., & Over, D. E. (2013). New paradigm psychology of reasoning: An introduction to the special issue edited by Elqayam, Bonnefon, and Over. *Thinking and Reasoning*, 19(3), 249-265.
- Evans, J. St. B. T. (2012). Questions ad challenges for the new psychology of reasoning. *Thinking and Reasoning*, 18(1), 5-31.
- Evans, J. St. B. T. , & Over, D. E. (2013). Reasoning to and from belief: Deduction and induction are still distinct. *Thinking and Reasoning*, 19(3), 267-283.
- Evans, J. St. B. T., Thompson, V., & Over, D. E. (2015). Uncertain deduction and conditional reasoning. *Frontiers in Psychology*, 6, 398.
- George, C. (1997). Reasoning from uncertain premises. *Thinking and Reasoning*, 3, 161-189.
- Gilio, A. (2002). Probabilistic reasoning under coherence in system P. *Annals of Mathematics and Artificial intelligence*, 34, 131-159.
- Gilio, A., & Over, D. P. (2012). The psychology of inferring conditionals from disjunctions : A probabilistic study. *Journal of Mathematical Psychology*, 56, 118-131.
- Hacking, I. (1975). The emergence of probability: A philosophy study of early ideas about probability, induction, and statistical inference. Cambridge: Cambridge University Press.

Hailperin, T. (1996). *Sentential probability logic*. Bethlehem: Lehigh University Press.

Hailperin, T. (2010). *Logic with a probability semantics*. Bethlehem: Lehigh University Press.

Jeffrey, R. (1991). Matter-of-fact conditionals. *The Aristotelian Society, Supplementary Vol. LXV*, 161-183.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove & London: Lawrence Erlbaum.

Mandel, D. R. (2014). The psychology of Bayesian reasoning. *Frontiers in Psychology*, 5, 1144.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57-74.

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

Oaksford, M., & Chater, N. (2009). Précis of *Bayesian rationality: The probabilistic approach to human reasoning*. *Behavioral and Brain Sciences*, 32, 69-84.

Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking and Reasoning*, 15(4), 431-438.

Mandel, D. R. (2014). The psychology of Bayesian reasoning, *Frontiers in Psychology*, 5, article 1144.

Pfeifer, N. (2014). Reasoning about uncertain conditionals. *Studia Logica*, 8, 1-18

Pfeifer, N., & Kleiter, G. D. (2005). Coherence and nonmonotonicity in human reasoning. *Synthese*, 146(1-2), 93-109.

Pfeifer, N., & Kleiter, G. D. (2006a). Inference in conditional probability logic. *Kybernetika*, 42(4), 391-404.

Pfeifer, N., & Kleiter, G. D. (2006b). Is human reasoning about nonmonotonic reasoning conditionals probabilistically coherent? *Proceedings of the 7th Workshop on Uncertainty Reasoning*, Mikulov (pp. 138-150).

- Pfeifer, N., & Kleiter, G. D. (2007). Human reasoning with imprecise probabilities: Modus ponens and denying the antecedent. *Fifth International Symposium on Imprecise Probability: Theories and applications* (pp. 347-356). Prague, 16-19 July 2007.
- Pfeifer, N., & Kleiter, G. D. (2009). Framing human inference by coherence based probability logic. *Journal of Applied Logic*, 7, 206-217.
- Pfeifer, N., & Kleiter, G. D. (2010). The conditional in mental probability logic. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals* (pp. 153-173). Oxford: Oxford University Press.
- Pfeifer, N., & Kleiter, G. D. (2011). Uncertain deductive reasoning. In K. Manktelow, D. Over & S. Elqayam (Eds.), *The science of reason* (pp. 145-166). Hove: Psychology Press.
- Politzer, G. (2015). Deductive reasoning under uncertainty: A Water tank analogy. (Working paper). Retrieved from http://jeannicod.ccsd.cnrs.fr/ijn_01140941
- Politzer, G. (in press). Deductive reasoning under uncertainty: A Water tank analogy. *Erkenntnis*.
- Politzer, G., Over, D. E., & Baratgin, J. (2010). Betting on conditionals. *Thinking and Reasoning*, 16(3), 172-197.
- Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, 90, 38-71.
- Rips, L. J., & Conrad, F. G. (1983). Individual differences in deduction. *Cognition and Brain Theory*, 6, 259-285.
- Singmann, H., Klauer, K. C., & Over, D. E. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in Psychology*, 5, article 316.
- Tentori, K., Crupi, V., & Russo, S. (2013). On the determinants of the conjunction fallacy: probability versus inductive confirmation. *Journal of Experimental Psychology: General*, 142(1), 235-255.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293-315.

Appendix

Table 2. Evaluation of the conclusion of six inference schemas. The columns show, for each given level of belief in the premise, the distributions of participants' choices (in percent) on the levels of belief in the conclusion (defined by comparison with the premise level). The responses that are coherent (for the uncertain premises) or valid (for the "sure" premises) are shaded. For each row, the top percentage refers to the *chips* scenario and the bottom percentage to the *activity* scenario.

Percent coherent: All the percentages differ significantly from random responding at the level of $p < .05$ unless indicated by °. The figures in parentheses refer to the number of participants in each group.

Table 2a. And-elimination.

AND-elimination		A AND C; therefore: A						
Evaluation of conclusion chances compared with premise chances		P r e m i s e c h a n c e s					0%	100%
		very low	low	average	high	very high		
greater	36	54	37	28	22	74	65	
	11	19	22	26	32	65		
greater or equal	11	12	33	11	26	21	12	
	11	15	4	22	7	12		
equal	14	19	26	28	33	6	79	
	59	63	70	43	46	23	83	
smaller or equal	4	0	4	28	11		4	
	4	0	0	0	0		0	
smaller	14	4	0	0	4		17	
	15	4	4	9	14		17	
smaller, equal, or greater	21	12	0	6	4			
	0	0	0	0	0			
		Mean						
Percent coherent	61°(28)	85(26)	96(27)	67°(18)	81(27)	79	100(34)	79(29)
	81(27)	97(27)	96(23)	91(23)	87(28)	89	100(26)	83(29)

Table 2b. And-introduction.

AND-introduction		A; C; therefore: A AND C						
Evaluation of conclusion chances compared with premise chances		P r e m i s e c h a n c e s					0%	100%
		very low	low	average	high	very high		
greater	13	3	7	14	31	21	10	
	9	13	0	16	23	10		
greater or equal	0	3	4	7	15	4	3	
	0	0	0	4	13	3		
equal	25	31	41	48	35	75	93	
	39	50	44	68	39	87	94	
smaller or equal	21	7	7	7	15		0	
	0	3	11	8	10		0	
smaller	42	55	37	24	4		7	
	52	30	41	4	16		6	
smaller, equal, or greater	0	0	4	0	0			
	0	0	3	4	0			
		Mean						
Percent coherent	88(24)	93(29)	85(27)	79(29)	54°(26)	80	75(28)	93(30)
	91(23)	83(30)	96(27)	80(25)	65°(31)	85	87(31)	94(31)

Table 2c. Or-introduction.

OR-introduction		A; therefore: A OR C							
Estimation of conclusion chances compared with premise chances		P r e m i s e c h a n c e s							
		very low	low	average	high	very high	0%	100%	
greater		33 18	44 26	41 52	62 37	57 26	70 62		
greater or equal		26 4	11 17	26 9	4 11	21 4	7 7		
equal		26 39	22 39	22 17	15 33	14 48	23 31	47 69	
smaller or equal		7 4	6 4	4 0	8 4	4 11		9 12	
smaller		0 36	17 13	7 17	4 15	4 11		44 19	
smaller, equal, or greater		7 0	0 0	0 4	8 0	0 0			
		Mean							
Percent coherent		85(27) 61°(28)	77(18) 82(23)	89(27) 78(23)	81(26) 81(27)	92(28) 78(27)	85 76	100(30) 100(29)	47°(34) 69(26)

Table 2d. And to If.

AND to IF		A AND C; therefore: IF A THEN C							
Estimation of conclusion chances compared with premise chances		P r e m i s e c h a n c e s							
		very low	low	average	high	very high	0%	100%	
greater		12 27	34 12	11 7	18 53	13 30	22 29		
greater or equal		12 0	3 12	15 19	18 10	21 0	7 6		
equal		42 43	34 60	48 48	50 30	50 57	70 65	83 55	
smaller or equal		12 0	14 0	7 0	0 0	4 0		10 6	
smaller		12 30	14 16	11 26	11 7	13 13		7 39	
smaller, equal, or greater		12 0	0 0	7 0	4 0	0 0			
		Mean							
Percent coherent		65°(26) 70°(30)	72(29) 84(25)	74(27) 74(27)	86(28) 93(30)	84(24) 87(23)	74 82	100(27) 100(31)	83(30) 55°(31)

Table 2e. Or to If-not.

OR to IF-NOT		A OR C; therefore: IF NOT-A THEN C							
Estimation of conclusion chances compared with premise chances		P r e m i s e c h a n c e s					0%	100%	
		very low	low	average	high	very high			
greater		23	10	22	10	17	37		
		29	8	15	10	9	42		
greater or equal		0	7	4	0	13	10		
		0	4	0	7	4	3		
equal		50	45	56	59	50	53	48	
		39	64	48	53	61	55	39	
smaller or equal		8	3	0	0	4		15	
		0	0	4	3	0		13	
smaller		8	24	15	31	17		37	
		32	20	33	20	26		48	
smaller, equal, or greater		12	10	4	0	0			
		0	4	0	7	0			
		Mean							
Percent coherent		66°(26)	72(29)	71(27)	90(29)	71(24)	76	53°(30)	100(27)
		71°(31)	84(25)	85(27)	76(30)	87(23)	81	55°(31)	100(31)

Table 2f. Contraposition.

Contraposition		IF A THEN C; therefore: IF NOT-C THEN NOT-A							
Estimation of conclusion chances compared with premise chances		P r e m i s e c h a n c e s					0%	100%	
		very low	low	average	high	very high			
greater		22	22	0	4	11	71		
		25	13	22	11	7	58		
greater or equal		0	6	4	4	6	12		
		4	0	0	7	4	4		
equal		37	22	41	52	18	18	13	
		32	43	39	44	22	38	28	
smaller or equal		0	0	11	4	11		10	
		4	0	4	7	7		3	
smaller		4	28	26	20	36		77	
		32	30	35	30	59		69	
smaller, equal, or greater		37	22	19	16	18			
		4	13	0	0	0			
		Mean							
Percent coherent		100(27)	100(18)	100(27)	100(25)	100(28)	100	100(34)	100 (30)
		100(28)	100(23)	100(23)	100(27)	100(27)	100	100(26)	100 (29)

Notes

1. The conditioning principle (Hacking 1975) assumes that the revised probability upon learning the outcome D at time t_1 is equal to the probability of H conditioned on the (imagined or assumed) evidence D at a moment t_0 (that is, $P(H|D)$ yielded by Bayes' rule) : $P_D(H) = P(H|D)$.
2. There is another possible methodology based on the notion of p-validity. An argument is p-valid when *the uncertainty of its conclusion cannot exceed the sum of the uncertainties of its premises* (Adams, 1998). The experimenter examines the participant's assessment of the argument's conclusion with reference to its formal validity. The two notions are not equivalent as a p-valid assessment can be incoherent. Some paradoxes may arise with the p-validity methodology. For example, it suffices that the sum of the uncertainties of the premises reaches 1 for the conclusion to be p-valid. For a comparison between coherence and p-validity, see Baratgin & Politzer (in press).
3. Participants who selected the *smaller* option in response to the *high* or *very high* probability of the premises of *AND-introduction* could do so for correct reasons with respect to the upper bound (rejecting an increase in probability and avoiding the classic conjunction fallacy) but also because they had an evaluation below the lower bound and so be incoherent. However, the small frequencies observed compared with the *very low*, *low* and *average* cases where the lower bound reaches zero suggests that such violations were rare, if they existed at all.
4. We owe this suggestion to N. Cruz.

Acknowledgments

The authors thank Nicole Cruz and David Over for very helpful comments on a previous draft of the manuscript.