

# Human reasoning with imprecise probabilities: Modus ponens and denying the antecedent

Niki Pfeifer

Department of Psychology,  
University of Salzburg,  
Austria  
niki.pfeifer@sbg.ac.at

Gernot D. Kleiter

Department of Psychology,  
University of Salzburg,  
Austria  
gernot.kleiter@sbg.ac.at

## Abstract

The MODUS PONENS ( $A \rightarrow B, A \therefore B$ ) is, along with MODUS TOLLENS and the two logically not valid counterparts DENYING THE ANTECEDENT ( $A \rightarrow B, \neg A \therefore \neg B$ ) and AFFIRMING THE CONSEQUENT, the argument form that was most often investigated in the psychology of human reasoning. The present contribution reports the results of three experiments on the probabilistic versions of MODUS PONENS and DENYING THE ANTECEDENT. In probability logic these arguments lead to conclusions with imprecise probabilities.

In the MODUS PONENS tasks the participants inferred probabilities that agreed much better with the coherent normative values than in the DENYING THE ANTECEDENT tasks, a result that mirrors results found with the classical argument versions. For MODUS PONENS a surprisingly high number of lower and upper probabilities agreed perfectly with the conjugacy property (upper probabilities equal one complements of the lower probabilities). When the probabilities of the premises are imprecise the participants do not ignore irrelevant (“silent”) boundary probabilities. The results show that human *mental probability logic* is close to predictions derived from *probability logic* for the most elementary argument form, but has considerable difficulties with the more complex forms involving negations.

**Keywords.** Mental probability logic, modus ponens, coherence, imprecise probabilities

## 1 Introduction

While there is a long tradition of probabilistic approaches in human judgment and decision making [10], only recently probabilistic approaches are adopted in the psychology of reasoning [16, 17, 7, 19, 12, 14, 18, 27, 26, 29]. Traditionally, classical logic dominated the psychology of human reasoning [6, 28, 2]. Classical logic was the normative standard

	MP	CMP	DA	CDA
$P_1:$	$A \rightarrow B$	$A \rightarrow B$	$A \rightarrow B$	$A \rightarrow B$
$P_2:$	$A$	$A$	$\neg A$	$\neg A$
<i>Concl.:</i>	$B$	$\neg B$	$\neg B$	$B$
<i>L-valid:</i>	yes	no	no	no
$V_i(\mathfrak{C})$	$t$	$f$	?	?

Table 1: Non-probabilistic version of the MODUS PONENS (MP), DENYING THE ANTECEDENT (DA), and their respective complementary versions (CMP, and CDA).  $A$  and  $B$  denote propositions.  $\rightarrow$  and  $\neg$  denote the material implication and negation, respectively, and are defined as usual. *L-valid* denotes logical validity, and  $V_i$  denotes the logical valuation-function  $V$  of the conclusion  $\mathfrak{C}$  under the interpretation  $i$  that assigns  $t$  (“true”), to all premises ( $P_1$  and  $P_2$ ). If the antecedents,  $A$ , of the conditional premise is false, then the truth value of the conclusion is not determined (denoted by the question mark).

of reference and used as a criterion for the rationality of human inferences. See Table 1 for often investigated argument forms in psychology.

Traditional psychological research on human reasoning designates human inference as rational/not rational if it corresponds to logically valid/not valid argument forms. Everyday life situations, though, are inherently uncertain. The uncertainty cannot be captured by classical logic. Reasoning about uncertainty and uncertain knowledge is a fundamental human competence. Thus, classical logic cannot be an adequate normative standard of reference for the psychology of reasoning.

We proposed a psychological theory of human reasoning, called “*mental probability logic*” [20, 22, 21, 23, 24, 25], which evaluates the rationality of human reasoning not by means of logical validity but by means of *coherence* [9, 8, 4]. *Mental probability logic* is a psychological *competence theory* about how humans

interpret common sense conditionals, represent the premises of everyday life arguments and draw inferences by coherent manipulations of mental representations.

Why a *competence* theory? Many investigations on cognitive processes report errors, fallacies, or biases. Well known are perceptual illusions, biases in judgment under uncertainty, or errors in deductive reasoning. While these phenomena may be startling and stimulating in the scientific process, they do not lead to theories that explain human performance in a systematic way. Collecting slips of the tongue does not lead to a theory of speaking. Psycholinguistics distinguishes performance and competence. Competence describes what functions a cognitive system can compute. Human reasoning can solve complex problems and perform sophisticated inferences. While developing a theory of reasoning one should have the explanation of these processes in mind. One should strive for a competence theory. The distinction between competence and performance was introduced by Noam Chomsky [3]. The analogy to deductive reasoning is obvious. The emphasis on the function a cognitive system should compute is due to David Marr [15].

*Mental probability logic* claims that the common sense conditionals are represented as subjective conditional probabilities. Based on the available information, the premises are evaluated and represented by coherent precise (point) probabilities, coherent imprecise probabilities, or logical information. Coherent imprecise probabilities can be represented by coherent interval probabilities or second order probability distributions. Human reasoning is a mental process that forms new representations from old ones by using probabilistic versions of formal inference rules. We assume that a certain core set of probabilistic inference rules are hard wired in the human inference engine.

The normative standard of *mental probability logic* is based on *coherence*. Coherence is the key concept in the tradition of subjective probability theory. It was originally developed by de Finetti [5]. More recent work includes [31, 13, 4, 8]. A probability assessment is coherent<sup>1</sup> if it does not admit one or more bets with sure loss. Coherence provides an adequate normative foundation for the *mental probability logic* and has many psychologically plausible advantages compared with classical concepts of probability:

- In the framework of coherence a *complete Boolean algebra* is not required for probabilistic inference. Full algebras are psychologically unrealistic as they can neither be unfolded in working

<sup>1</sup>Throughout we use “coherent” as synonymous with “totally coherent” [9].

memory nor be stored in the long term memory. *Mental probability logic* suggests that humans try to keep the memory load as small as possible and process only relevant informations (see also [30]).

- *Conditional probability*,  $P(B|A)$ , is a *primitive* notion. The probability values are assigned *directly*. Conditional probability is not “defined”—as in probability textbooks—via the fraction of the “joint”,  $P(A \wedge B)$ , and the “marginal”,  $P(A)$ , probabilities.<sup>2</sup> Conditional probabilities are *directly* encoded or just directly connected to the arguments of the if-then relation.
- Because lack of knowledge (time, effort) it may be impossible for a person to assign precise probabilities to an event. If a person is uncertain about probabilities, then *mental probability logic* supposes that human subjects make coherent *imprecise* probabilistic assessments (by interval-valued probabilities or by second order probability distributions).
- Coherence is in the tradition of subjective probability theory in which probabilities are conceived as *degrees of belief*. Degrees of belief are naturally affine to psychology.

Imprecise versions of the argument forms presented in Table 1 are formalized by interval probabilities [23] or by second order probability distributions [25]. In the present study we focus on interval probabilities only. We now present imprecise versions of the four argument forms of Table 1. While only the MODUS PONENS is logically valid, all four argument forms admit to infer coherent probability intervals for the conclusion.

The imprecise version of the MODUS PONENS has the form:

$$\begin{aligned} P(B|A) &\in [x', x''], P(A) \in [y', y''] \\ \therefore P(B) &\in [x'y', 1 - y' + x''y']. \end{aligned} \quad (1)$$

Since  $P(\neg\mathcal{C}) = 1 - P(\mathcal{C})$  (conjugacy principle [31]) the complement of an interval  $[l, u]$  is  $[1 - u, 1 - l]$ , it trivially follows that the imprecise COMPLEMENT MODUS PONENS has the form:

$$\begin{aligned} P(B|A) &\in [x', x''], P(A) \in [y', y''] \\ \therefore P(\neg B) &\in [y' - x''y', 1 - x'y']. \end{aligned} \quad (2)$$

The imprecise DENYING THE ANTECEDENT has the form:

$$\begin{aligned} P(B|A) &\in [x', x''], P(\neg A) \in [y', y''] \\ \therefore P(\neg B) &\in [(1 - x'')(1 - y''), 1 - x'(1 - y'')] \end{aligned} \quad (3)$$

<sup>2</sup>The definition  $P(B|A) =_{df.} P(A \wedge B)/P(A)$  is problematic if  $P(A) = 0$ .

The imprecise COMPLEMENT DENYING THE ANTECEDENT has the form:

$$P(B|A) \in [x', x''], P(\neg A) \in [y', y''] \\ \therefore P(B) \in [x'(1 - y''), x'' + y'' - x''y''] \quad (4)$$

Equations (1)-(4) may be obtained by natural extension [31], likewise, by de Finetti's Fundamental Theorem [5] or by Lad's generalized version [13]—or by elementary probability theory (for a demonstration see [23]).

In the psychological literature, the non-probabilistic versions of the MODUS PONENS and the DENYING THE ANTECEDENT were studied extensively. Meta-analytical results show that the MODUS PONENS is endorsed by 89-100% of human subjects [6]. The DENYING THE ANTECEDENT is endorsed by 17-73% of the subjects [6]. We do not judge the 17-73% of subjects as irrational. Rather, we propose to reinterpret the data in the light of *mental probability logic*. The question is not whether the human subjects endorse the non-probabilistic DENYING THE ANTECEDENT, but whether they infer coherent probabilities from the premises of the *imprecise* DENYING THE ANTECEDENT. In the next sections we present empirical data on the four imprecise argument forms (1)-(4).

## 2 Experiment 1

### 2.1 Method and Procedure

Thirty students of the University of Salzburg participated in Experiment 1. No students with special logical or mathematical education were included.

Each participant received a booklet containing a general introduction, one example explaining the response modality with point percentages, and one example explaining the response modality with interval percentages. Three target tasks were presented on separate pages. Eight additional target tasks were presented in tabular form. The first three MODUS PONENS target tasks had the following form:

Please imagine the following situation. Several cars are parked on a parking lot. About these cars we know the following:

*Exactly 80% of the red cars on this parking lot are two-door-cars.*

*Exactly 90% of the cars on this parking lot are red cars.*

Imagine all the cars that are on this parking lot. How many of these cars are *two-door-cars*?

Then, the participants were informed that the solu-

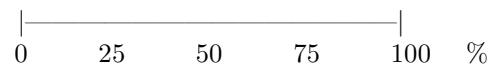
tion is either a point percentage or a percentage between two boundaries (*from at least ... to at most ...*). The booklet offered two response modalities where the participants had to choose one deliberately.

#### Response Modality 1:

If you think that the correct answer is a *point* percentage, please fill in your answer here:

*Exactly* .....% of the cars on this parking lot are *two-door-cars*.

**Point percentage:**

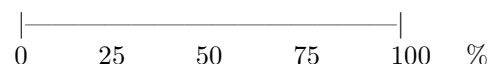


#### Response Modality 2:

If you think that the correct answer lies within two boundaries (*from at least ... to at most ...*), please mark the two values here:

*At least* .....% and *at most* .....% of the cars on this parking lot are *two-door-cars*.

**Within the bounds of:**



The subsequent two tasks were formulated accordingly. In the second task the numerical values in two premises were 20 and 40%, and in the third task 60 and 90%, respectively. Each task was on a separate page. After the third task the participants answered eight analogous tasks presented on one page in tabular form. Again, the cover story was kept constant, only the numerical values contained in the premises varied (see Table 2).

The thirty participants were divided into two groups, fifteen participants received the MODUS PONENS tasks, as just described, and fifteen participants received DENYING THE ANTECEDENT tasks. The DENYING THE ANTECEDENT tasks were formulated exactly as the MODUS PONENS tasks, with two differences. First, a negation was added in the second premise: "*Exactly 90% of the cars on this parking lot are **not** red cars*". Second, a negation was added to the question: "How many of these cars are ***not** two-door-cars*?". Both response modalities were adopted accordingly. Adding the negations to the second premise and to the conclusion as just described clearly reflects the form of the corresponding imprecise DENYING THE ANTECEDENT. In both conditions, we presented the same percentage numbers to the participants.

The booklets were mixed and assigned arbitrarily to

$P_1$	$P_2$	CLB	CUB	LBR	UBR
80	90	72	82	75.80 (5.16)	82.60 (8.33)
20	40	8	68	14.60 (11.15)	55.60 (30.24)
60	90	54	64	51.93 (10.13)	63.87 (14.21)
40	40	16	76	25.13 (12.65)	64.07 (35.43)
80	70	56	86	53.00 (21.91)	71.20 (27.26)
20	60	12	52	18.60 (13.37)	46.27 (23.59)
100	100	100	100	100.00 (0.00)	100.00 (0.00)
60	70	42	72	48.00 (15.08)	67.60 (22.04)
40	60	24	64	33.27 (13.88)	59.47 (23.55)
70	80	56	76	61.33 (13.80)	76.73 (12.40)
30	50	15	65	20.67 (12.80)	51.33 (30.32)

Table 2: Mean lower (LBR) and mean upper bound responses (UBR) of the MODUS PONENS tasks of Experiment 1 ( $n_1 = 15$ ). The standard deviations are in parenthesis.  $P_1$  and  $P_2$  denote the percentages in the premises. CLB and CUB denote the normative/coherent lower and upper bounds, respectively.

the participants. All participants were tested individually in a quiet test room in the department. They were told to take as much time as they wanted. In case of questions, they were asked to reread the instructions carefully.

## 2.2 Results and Discussion

Table 2 lists the probabilities presented in the premises, the normative lower and upper bounds, and the participants’ mean lower and upper bound responses for the MODUS PONENS tasks. In the task with certain premises (i.e., “100%” in both premises) all fifteen participants responded with point value of 100%, which is normatively correct.

Table 3 lists the respondents’ mean lower and upper bound responses for the DENYING THE ANTECEDENT tasks. In the task with certain premises four of the fifteen participants responded with the unit interval 0-100%. These four participants clearly understood that the DENYING THE ANTECEDENT is probabilistically not informative if all premises have probabilities equal to 1. One participant responded with the point value 100 and one with the point value 50%. The majority (nine out of the fifteen participants) responded with the point value 0%.

In the MODUS PONENS tasks on the average 30% of the responses were point values (the task with the certain premises not included). In the DENYING THE ANTECEDENT tasks on the average 23% of the responses were point values (the task with the certain premises was not averaged).

In both conditions the standard deviations are high.

$P_1$	$P_2$	CLB	CUB	LBR	UBR
80	90	2	92	27.13 (33.58)	64.13 (37.94)
20	40	48	88	36.07 (27.71)	62.80 (31.97)
60	90	4	94	22.47 (22.17)	73.00 (29.30)
40	40	36	76	33.87 (20.87)	62.67 (23.25)
80	70	6	76	22.93 (23.20)	55.27 (34.74)
20	60	32	92	29.07 (17.90)	66.53 (24.65)
100	100	0.00	100	10.00 (28.03)	36.67 (48.06)
60	70	12	82	28.40 (26.26)	60.93 (30.39)
40	60	24	84	22.53 (17.01)	59.07 (23.71)
70	80	6	86	19.93 (23.91)	54.93 (31.25)
30	50	35	85	27.87 (20.76)	62.67 (25.13)

Table 3: Mean lower LBR and mean upper bound responses UBR of the DENYING THE ANTECEDENT tasks of Experiment 1 ( $n_2 = 15$ ). The standard deviations are in parenthesis.  $P_1$  and  $P_2$  denote the percentages in the premises. CLB and CUB denote the normative/coherent lower and upper bounds, respectively.

This may be a consequence of the explicit presentation of the point value response modality. The participants could actually have some imprecise value in mind, but nevertheless respond with a representative point value just to reduce the complexity of the task. Such point value responses bias the mean lower and upper bound responses. To avoid constantly pointing explicitly to the possibility to give an interval value response we dropped the response point response modality in Experiment 2. Dropping the point response modality forces the participants to respond by intervals while still allowing point value responses by equating the lower and the upper bound responses.

Table 4 reports the frequencies of interval response categories of the MODUS PONENS condition in  $3 \times 3$  tables. Each table contains the six possible interval responses together with the according empirical frequencies of the interval responses. The *columns* designate whether the participants’ lower bounds are below (*LB*), within (*LW*), or above (*LA*) the normative intervals. The *rows* designate whether the upper bounds are above (*UA*), within (*UW*) or below (*UB*) the normative intervals.

Table 5 reports the frequencies of interval response categories of the DENYING THE ANTECEDENT condition in  $3 \times 3$  tables. Figure 1 presents the averaged interval response frequencies in the MODUS PONENS tasks. The data of Task 2 and of Task 7 were not averaged. The normative lower bound of Task 2 is  $\leq 10\%$ . Both normative bounds of task 7 are equal to 100%. Figure 1 shows that in the MODUS PONENS tasks the majority of the participants gave coherent interval responses.

Figure 2 presents the averaged interval response fre-

quencies in the DENYING THE ANTECEDENT tasks. The data of Task 1, 3, 5, 6, 7, and 10 were not averaged. The normative upper bounds of Task 1, 3, 6, and 7 are  $\geq 90\%$ . The normative lower bounds of Task 2, 5, 7 and 10 are  $\leq 10\%$ . Figure 2 shows that in the DENYING THE ANTECEDENT tasks the participants responded with more incoherent intervals than in the MODUS PONENS tasks. More coherent interval responses were observed in the MODUS PONENS tasks (62.93% of the participants) compared with the DENYING THE ANTECEDENT tasks (41.33% of the participants).

In the MODUS PONENS condition, the mean responses agree very well with the normative lower ( $r_{LBR,CLB} = .99$ ) and upper ( $r_{UBR,CUB} = .92$ ) probabilities. The good agreement remains when the lower ( $r_{(LBR,CLB).P1} = .91$ ) and upper ( $r_{(UBR,CUB).P2} = .95$ ) percentages in the premises are partialled out. Partialling out the values contained in the tasks reduces the possible influence of anchoring and/or matching effects.

In the DENYING THE ANTECEDENT condition we observed a different pattern. While the mean responses still agree well with the normative lower ( $r_{(LBR,CLB)} = .76$ ) probabilities, the correlation is slightly negative for the upper ( $r_{(UBR,CUB)} = -.20$ ) probabilities. Partialling out Premise 1 and Premise 2 reduces the correlations to  $r_{(LBR,CLB).P1} = .25$  and  $r_{(UBR,CUB).P2} = .03$ . The results may be explained by assuming that the participants just respond with values close to those contained in the description of the tasks (known as “matching heuristic”).

It is well known that logical tasks involving negations are difficult. In probabilistic inference tasks we consider the correlations between the probabilities of the premises and the normative lower and upper probabilities of the conclusions. For the set of our MODUS PONENS tasks the four correlations are all positive,  $r_{(P1,CLB)} = .97$  and  $r_{(P2,CLB)} = .92$  for the lower probabilities, and  $r_{(P1,CUB)} = .86$  and  $r_{(P2,CUB)} = .51$  for the upper ones.

For the DENYING THE ANTECEDENT tasks (with the identical numerical probabilities of the premises!) the lower bound correlations are highly negative,  $r_{(P1,CLB)} = -.92$  and  $r_{(P2,CLB)} = -.93$ , and for the upper bounds positive,  $r_{(P1,CUB)} = .24$  and  $r_{(P2,CUB)} = .62$ . The weighting and integration of affirmative and non-affirmative information makes tasks like the DENYING THE ANTECEDENT especially difficult. We note that linear regression predicts lower probabilities better than upper ones.

The overall conclusion of Experiment 1 is that the responses of the participants in the MODUS PONENS con-

	Schema			Task 1			Task 2		
UA	<i>a</i>	<i>b</i>	<i>c</i>	0	4	1	0	4	0
UW	<i>d</i>	<b><i>e</i></b>	-	0	<b>10</b>	-	0	<b>11</b>	-
UB	<i>f</i>	-	-	0	-	-	0	-	-
	LB	LW	LA	LB	LW	LA	LB	LW	LA
	Task 3			Task 4			Task 5		
UA	0	5	0	0	5	0	0	2	0
UW	2	<b>7</b>	-	0	<b>9</b>	-	1	<b>10</b>	-
UB	1	-	-	1	-	-	2	-	-
	LB	LW	LA	LB	LW	LA	LB	LW	LA
	Task 6			Task 7			Task 8		
UA	0	4	1	-	-	-	0	4	1
UW	1	<b>9</b>	-	0	<b>15</b>	-	0	<b>9</b>	-
UB	0	-	-	0	-	-	1	-	-
	LB	LW	LA	LB	LW	LA	LB	LW	LA
	Task 9			Task 10			Task 11		
UA	0	5	0	0	3	1	0	4	0
UW	0	<b>10</b>	-	1	<b>10</b>	-	0	<b>11</b>	-
UB	0	-	-	0	-	-	0	-	-
	LB	LW	LA	LB	LW	LA	LB	LW	LA

Table 4: Frequencies of the interval responses in the MODUS PONENS condition of Experiment 1 ( $n_1 = 15$ ). UA: the participants’ upper bound response is *above* the normative upper bound, UW: upper bound response is *within* the normative interval, UB: upper bound response is *below* the normative lower bound; LA, LW, and LB: same for the participants’ lower bound responses. *a*: too wide interval responses, *b*: lower bound responses coherent, *c*: both bound responses above, *d*: upper bound responses coherent, *e*: both bound responses coherently within  $\pm 5\%$  (bold), *f*: both bound responses below the normative lower bounds.

dition are very close to the normative values while in the DENYING THE ANTECEDENT condition the responses might be explained by matching based guessing.

The presence of the negations in the DENYING THE ANTECEDENT is a possible explanation, why there were less coherent interval responses compared with the MODUS PONENS tasks. It is easier to cognitively represent an affirmed than a negated proposition. An affirmed proposition can be visualized, for example, more directly than a negated one. Classical MODUS PONENS was proposed as a basic and “hard wired” inference rule [28, 2]. Probabilistic MODUS PONENS is a similar candidate.

In addition to the MODUS PONENS and the DENYING THE ANTECEDENT, the respective complementary versions are investigated in Experiment 2. By investigating the complementary versions as well, the presence of the negation is more balanced.

	Schema			Task 1			Task 2		
UA	a	b	c	-	2	0	0	1	0
UW	d	e	-	-	13	-	3	6	-
UB	f	-	-	-	-	-	5	-	-
	LB	LW	LA	LB	LW	LA	LB	LW	LA
	Task 3			Task 4			Task 5		
UA	0	1	0	0	3	0	0	4	0
UW	0	14	-	5	6	-	2	9	-
UB	0	-	-	1	-	-	0	-	-
	LB	LW	LA	LB	LW	LA	LB	LW	LA
	Task 6			Task 7			Task 8		
UA	0	2	0	-	-	-	0	4	0
UW	6	6	-	-	15	-	4	7	-
UB	1	-	-	-	-	-	8	-	-
	LB	LW	LA	LB	LW	LA	LB	LW	LA
	Task 9			Task 10			Task 11		
UA	0	2	0	0	2	0	0	2	0
UW	5	7	-	3	10	-	6	5	-
UB	1	-	-	0	-	-	2	-	-
	LB	LW	LA	LB	LW	LA	LB	LW	LA

Table 5: Frequencies of the interval responses in the DENYING THE ANTECEDENT condition ( $n_2 = 15$ ). For explanation of the schema see Table 4.

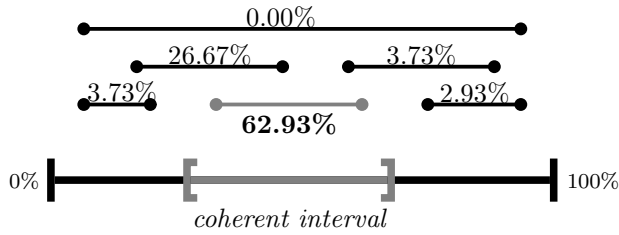


Figure 1: Averaged interval response frequencies over nine selected MODUS PONENS tasks (see text,  $n_1 = 15$ ).

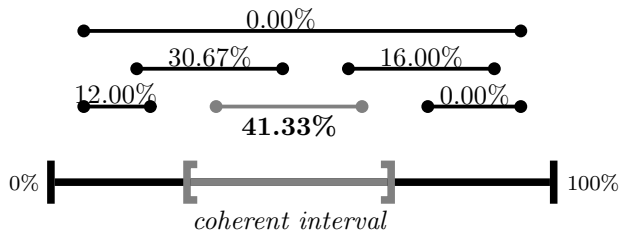


Figure 2: Averaged interval response frequencies over five selected DENYING THE ANTECEDENT tasks (see text,  $n_2 = 15$ ).

## 3 Experiment 2

### 3.1 Method and Procedure

Method and procedure of Experiment 2 are analog to Experiment 1. Sixty students of the University of Salzburg participated in Experiment 2. No students with special logical or mathematical education were included. Thirty participants were assigned to the MODUS PONENS condition and thirty participants were assigned to the DENYING THE ANTECEDENT condition.

In the MODUS PONENS condition, each participant worked out three MODUS PONENS tasks and three COMPLEMENT MODUS PONENS tasks. To counterbalance position position effects, fifteen participants got the MODUS PONENS tasks at the beginning, and fifteen participants got the MODUS PONENS tasks at the end of the session. The MODUS PONENS tasks had the following form:

Please imagine the following situation. Around Christmas time a certain ski-resort is very busy. This region is very popular among sportsmen, like skiers, snow-boarders, and sledge-rider. Every hour a cable-car brings the sportsmen to the top. About this cable-car we know:

*Exactly 100% of the skiers wear red caps.*  
*Exactly 100% of the sportsmen are skiers.*

Imagine all the sportsmen in this cable car. How many of these sportsmen wear a *red cap*?

Speaking about a closed room (cable-car) instead of an unspecified parking lot (Experiment 1) should help to represent and visualize the problems. As in Experiment 1, participants were free to respond either in terms of point percentages or in terms of interval percentages. In Experiment 2, however, the response modality 1 (point response) was dropped. The participants were informed by two examples at the beginning that point values can be given by equating the lower and the upper bounds.

All three MODUS PONENS tasks had the same structure. The percentages of the two premises in the first task were 100 and 100%, in the second task were 70 and 90%, and in the third task the percentages were 70 and 50%, respectively. The three COMPLEMENT MODUS PONENS tasks contained the same percentages and differed from the MODUS PONENS task only in one respect: a negation was added to the conclusion (“How many of these sportsman **do not** wear a *red cap*?”).

The DENYING THE ANTECEDENT condition was analogue. Fifteen participants first received the three DENYING THE ANTECEDENT tasks and then the three complementary versions of the DENYING THE AN-

$P_1$	$P_2$	CLB	CUB	LBR	UBR
MODUS PONENS					
100	100	100	100	100.00 (.00)	100.00 (.00)
70	90	63	73	62.43 (11.77)	69.17 (9.71)
70	50	35	85	42.5 (15.13)	54.83 (21.19)
COMPLEMENT MODUS PONENS					
100	100	.00	.00	.00 (.00)	.00 (.00)
70	90	27	37	35.40 (16.73)	42.03 (17.45)
70	50	15	65	41.00 (18.82)	53.67 (17.71)
DENYING THE ANTECEDENT					
100	100	.00	100	37.37 (47.53)	85.00 (35.11)
70	20	20	44	18.63 (15.25)	41.63 (15.97)
70	50	15	65	25.4 (21.12)	59.23 (20.73)
COMPL. DENYING THE ANTECEDENT					
100	100	.00	100	0.83 (4.56)	53.33 (49.01)
70	20	56	80	51.9 (19.12)	75.87 (20.19)
70	50	35	85	32.70 (12.92)	65.17 (27.43)

Table 6: Mean lower LBR and mean upper bound responses UBR of the MODUS PONENS condition ( $n_1 = 30$ ) and of the DENYING THE ANTECEDENT condition ( $n_2 = 30$ ) of Experiment 2. The standard deviations are in parenthesis.  $P_1$  and  $P_2$  denote the percentages in the premises. CLB and CUB denote the normative/coherent bounds.

TECEDENT tasks. The order was reversed for the other fifteen participants. The premises and the conclusions were adopted accordingly. The only difference in the percentages between the MODUS PONENS condition and the DENYING THE ANTECEDENT condition was, that “90%” was replaced by “20%”. The reason for this was to avoid non-informative assessments.

### 3.2 Results and Discussion

The results of t-tests indicate that there were no position effects. We therefore pooled the data in the MODUS PONENS condition ( $n_1 = 30$ ) and in the DENYING THE ANTECEDENT condition ( $n_2 = 30$ ).

Table 6 lists the probabilities presented in the premises, the normative lower and upper bounds, and the participants’ mean lower and upper bound responses for the MODUS PONENS tasks and the COMPLEMENT MODUS PONENS tasks of Experiment 2.

In the MODUS PONENS tasks with certain premises (“100%” in both premises) all thirty participants responded with that point value 100%, which is normatively correct. In the according COMPLEMENT MODUS PONENS tasks all thirty participants responded correctly with the point value 0.00%. Thus, in Task 1 all participants inferred (correctly) point values. In the other tasks (both MODUS PONENS and COMPLEMENT

MODUS PONENS), between 50 and 60% of the responses were point value responses, which is about double compared with Experiment 1.

This result is surprising, since dropping the explicit point value response modality should decrease and not increase the number of point value responses. A possible explanation is that, as for all participants the first task contained certain premises and as all participants responded by point values in the first task, they simply continued to give point value responses later on.

In the DENYING THE ANTECEDENT tasks with certain premises (i.e., “100%” in both premises) fourteen of the thirty participants inferred a unit interval,  $[\leq 1, 100]\%$ . Four participants inferred a point value equal to zero, and ten inferred a point value equal to 100%. One participant inferred a point value of 50% and one an interval between 70 and 100%. In the according COMPLEMENT DENYING THE ANTECEDENT tasks fifteen of the thirty participants inferred a unit interval,  $[\leq 1, 100]\%$ . Thirteen responded a point value equal to zero. One participant inferred a  $[25, 50]\%$  interval and one inferred a  $[0, 50]\%$  interval. Practically half of the participants understood that only a non-informative interval can be inferred if each premise is certain.

In the two DENYING THE ANTECEDENT tasks with 70 and 20%, and 70 and 50%, in the premises, 0 and 16.67% point value responses were observed, respectively. This amount of point value responses is smaller than in Experiment 1. In both according COMPLEMENT DENYING THE ANTECEDENT tasks 26.67% point value responses were observed, which is comparable to the results of Experiment 1.

Table 7 reports the frequencies of interval response categories of the MODUS PONENS condition and of the DENYING THE ANTECEDENT condition in  $3 \times 3$  tables. 80.22% of the participants inferred coherent intervals in the MODUS PONENS condition on the average (the tasks with certain premises were not averaged). 55.00% of the participants inferred coherent intervals in the DENYING THE ANTECEDENT condition on the average (the tasks with certain premises were not averaged).

In the MODUS PONENS tasks 85.00% (Experiment 1: 62.93%) of the interval responses were coherent on the average. In the DENYING THE ANTECEDENT tasks 56.66% (Experiment 1: 41.33%) of the interval responses were coherent on the average. The improved cover-story explains why more coherent interval responses in Experiment 2 than in Experiment 1 were observed.

MODUS PONENS tasks							
	Task 2			Task 3			Coh. Bounds
UA	0	4	1	0	0	0	Task 2:
UW	2	<b>22</b>	-	1	<b>29</b>		63–73
UB	1	-	-	0	-	-	Task 3:
	LB	LW	LA	LB	LW	LA	35–85
COMPLEMENT MODUS PONENS tasks							
	Task 2			Task 3			Coh. Bounds
UA	0	0	6	0	2	1	Task 2:
UW	5	<b>19</b>	-	0	<b>27</b>		27–37
UB	0	-	-	0	-	-	Task 3:
	LB	LW	LA	LB	LW	LA	15–65
DENYING THE ANTECEDENT tasks							
	Task 2			Task 3			Coh. Bounds
UA	3	4	0	0	5	0	Task 2:
UW	6	<b>16</b>	-	7	<b>18</b>		20–44
UB	1	-	-	0	-	-	Task 3:
	LB	LW	LA	LB	LW	LA	15–65
COMPLEMENT DENYING THE ANTECEDENT tasks							
	Task 2			Task 3			Coh. Bounds
UA	1	8	0	1	6	0	Task 2:
UW	6	<b>14</b>	-	4	<b>18</b>		56–80
UB	1	-	-	1	-	-	Task 3:
	LB	LW	LA	LB	LW	LA	35–85

Table 7: Frequencies of the interval responses in the MODUS PONENS ( $n_1 = 30$ ) and the DENYING THE ANTECEDENT condition ( $n_2 = 30$ ) of Experiment 2. For explanation see Table 4.

All participants inferred a probability(interval) of a conclusion  $\mathcal{C}$ ,  $P(\mathcal{C}) \in [z'_{\mathcal{C}}, z''_{\mathcal{C}}]$ , and the probability of the negated conclusion,  $P(\neg\mathcal{C}) \in [z'_{\neg\mathcal{C}}, z''_{\neg\mathcal{C}}]$ . To test the conjugacy principle of the interval responses, we checked for each participant whether (i)  $z'_{\mathcal{C}} + z''_{\neg\mathcal{C}} = 100\%$ , and whether (ii)  $z'_{\neg\mathcal{C}} + z''_{\mathcal{C}} = 100\%$ .

In the MODUS PONENS tasks with certain premises, all participants satisfied both equalities, (i) and (ii). In the tasks with 70% and 90% in the premises sixteen of the thirty participants satisfied both, (i) and (ii). In the tasks with 70% and 50% in the premises fifteen of the thirty participants satisfied both, (i) and (ii). It is surprising that in the MODUS PONENS tasks more than half of the participants gave intervals that with lower/upper probabilities that exactly add up to 1. In the DENYING THE ANTECEDENT tasks with certain premises, twenty of the thirty participants satisfied both, (i) and (ii). In the tasks with 70 and 20% premises nobody satisfied both, (i) and (ii), eleven satisfied (i), and one satisfied (ii). In the tasks with 70 and 50% ten satisfied both, (i) and (ii).

In sum, more additive responses and more coherent interval responses were observed in MODUS PONENS tasks than in the DENYING THE ANTECEDENT tasks. It is

reasonable that humans are better in argument forms that guarantee high probabilities of the conclusion if each premise is highly probable. If the premises of the MODUS PONENS are certain, then the conclusion is certain. However, if the premises of the DENYING THE ANTECEDENT are certain, then the probability of the conclusion is in the unit interval  $[0, 1]$ .

## 4 Experiment 3

This section reports data of an experiment with imprecise probabilities in the premises conducted by Florian Bauerecker [1]. Specifically, we focus on human understanding of what we call “silent bounds”. We call a probability bound  $b$  of a premise *silent* if, and only if,  $b$  is *irrelevant* for the probability propagation from the premise(s) to the conclusion. E.g., in the probabilistic MODUS PONENS  $y''$  is silent ( $y''$  doesn’t occur in the lower or upper probabilities of the conclusion, see (1)). Experiment 3 introduces an especially critical test of the claim that human subjects are capable to make coherent probabilistic inferences.

Method and procedure of Experiment 3 are analog to Experiment 1. Eighty participants were recruited for investigating questions going beyond the scope of the present study. Therefore, we report selected data on the imprecise MODUS PONENS only ( $n = 40$ ). The MODUS PONENS tasks were formulated as follows:

Please imagine the following situation. Claudia works at blood donation services. She investigates to which blood group the donated blood belongs and whether the donated blood is Rhesus-positive.

Claudia is 60% certain: If the donated blood belongs to the blood group 0, then the donated blood is Rhesus-positive.

Claudia knows as well that donated blood belongs with *more than 75%* certainty to the blood group 0.

How certain should Claudia be that a recent donated blood is Rhesus-positive?

Contrary to Experiment 1 and Experiment 2, the conditional premise is here formulated in a if-then form. The cover-story remained constant, only the numbers in the premises varied. Table 8 lists the probabilities presented in the premises, the normative lower and upper bounds, and the participants’ mean lower and upper bound responses for the MODUS PONENS tasks with and without silent bounds.

The participants inferred higher upper bounds in the MODUS PONENS task containing silent bounds ( $M = 71.79$ ) compared with the according task not containing silent bounds ( $M = 60.20$ ;  $t(39)=3.53$ ,  $p=.001$ ).



$P_1/P_2$	CI	LBR	UBR
60/[75,100*]	[45,70]	44.50 (21.57)	71.78 (20.07)
60/75	[45,70]	46.83 (23.76)	60.20 (16.86)
[75,100]/60	[45,100]	43.42 (22.00)	72.38 (22.98)
75/60	[45,85]	46.27 (21.73)	59.90 (17.19)

Table 8: Mean lower LBR and mean upper bound responses UBR of the MODUS PONENS tasks ( $n = 40$ ) in [1]. \* denotes the silent bound. The standard deviations are in parenthesis.  $P_1$  and  $P_2$  denote the percentages in the premises. CI denotes the normative/coherent interval.

Thus the participants were sensitive to the silent bounds. They did not understand the irrelevance of the silent bound for the probability propagation from the premises to the conclusion.

[21] report data on a conjunction problem where in one condition interval-values in the premises were presented. All upper bounds were equal to 100%. In the other condition only corresponding point values were presented. The point values were equal to the lower bounds of the interval condition. Higher mean lower bounds were observed in the interval condition than in the point condition. An explanation for this finding is, that the participants reduced the processing load of the interval valued premises by representing only the means of the lower and upper bounds. Then, of course, the coherent lower bound must be higher.

This explanation of the [21] data on the conjunction problem is, however, not applicable to the data from the imprecise MODUS PONENS task. If the second premise (containing silent bounds) is represented as 88%, then the coherent interval of the conclusion is [53%, 65%]. Assuming that the participants represent “88” instead of the interval “[75, 100]”, then the participants’ mean upper bounds should be *lower* in the interval value condition than in the point value condition.

An alternative explanation is that higher explicit imprecision (by communicating interval-values in the premises) elicits larger interval responses. It could be that *conversational implicatures* [11] modulate the accumulation of imprecision. The participant assumes by conversational implicature that the experimenter communicates only relevant informations. Thus the silent bound is not understood as irrelevant, rather, the silent bound is understood by the participant as a hint from the experimenter to add imprecision to the conclusion: to infer wider intervals.

	Task 1 (60/[75,100*])			Task 2 (60/75)			Coh. Bounds
UA	1	14	0	0	2	4	Task 1:
UW	8	<b>16</b>	-	7	<b>25</b>		45–70
UB	1	-	-	2	-	-	Task 2:
	LB	LW	LA	LB	LW	LA	45–70
	Task 3 ([75,100]/60)			Task 4 (75/60)			Coh. Bounds
UA	-	-	-	0	2	0	Task 3:
UW	7	<b>32</b>	-	8	<b>29</b>		45–100
UB	1	-	-	1	-	-	Task 4:
	LB	LW	LA	LB	LW	LA	45–85

Table 9: Frequencies of the interval responses in the MODUS PONENS ( $n = 40$ ) tasks of Experiment 3. Coherent interval responses are bold ( $\pm 5\%$  tolerance interval). Further explanation in Table 4.

## 5 Concluding Remarks

We reported three psychological experiments on the probabilistic versions of two prominent argument forms in the framework of *mental probability logic*. Clearly more coherent responses were observed in MODUS PONENS than in DENYING THE ANTECEDENT tasks. Human subjects employ inference rules that guarantee high probability conclusions if the premises are highly probable. Practically no participant inferred “too wide” intervals such that the coherent intervals are subintervals. While most participants did not understand the irrelevance of the silent bounds in the MODUS PONENS task of Experiment 3 they are not completely “blind” for them. The close agreement of the mean responses and the normative values of the lower probabilities in Experiment 3 is stunning. One may speculate that human subjects are doing better in processing lower than upper probabilities. More than half of the participants responded with lower/upper probabilities that agreed perfectly with the conjugacy principle.

## References

- [1] F. Bauerecker. Konditionales Schließen mit Intervallen und Verständnis für das Komplement einer Schlussform. Master’s thesis, Universität Salzburg, 2006.
- [2] M. D. S. Braine and D. P. O’Brien, editors. *Mental logic*. Erlbaum, Mahwah, 1998.
- [3] N. Chomsky. *Aspects of the theory of syntax*. MIT Press, Cambridge, 1965.
- [4] G. Coletti and R. Scozzafava. *Probabilistic logic in a coherent setting*. Kluwer, Dordrecht, 2002.

- [5] B. De Finetti. *Theory of probability*, volume 1, 2. John Wiley & Sons, Chichester, 1974. Original work published 1970.
- [6] J. S. B. T. Evans, S. E. Newstead, and R. M. J. Byrne. *Human Reasoning*. Erlbaum, Hove, 1993.
- [7] J. S. B. T. Evans, S. H. Handley, and D. E. Over. Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29:321–355, 2003.
- [8] A. Gilio. Probabilistic reasoning under coherence in System P. *Annals of Mathematics and Artificial Intelligence*, 34:5–34, 2002.
- [9] A. Gilio and S. Ingrassia. Totally coherent set-valued probability assessments. *Kybernetika*, 34(1):3–15, 1998.
- [10] T. Gilovich, D. Griffin, and D. Kahneman, editors. *Heuristics and biases. The psychology of intuitive judgment*. Cambridge University Press, Cambridge, 2002.
- [11] H. P. Grice. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and semantics*, volume 3: Speech acts. Seminar Press, New York, 1975.
- [12] P. N. Johnson-Laird, V. Girotto, P. Legrenzi, and M. S. Legrenzi. Naive probability: A mental model theory of extensional reasoning. *Psychological Review*, 106(1):62–88, 1999.
- [13] F. Lad. *Operational subjective statistical methods: A mathematical, philosophical, and historical introduction*. Wiley, New York, 1996.
- [14] I.-M. Liu, K.-C. Lo, and J.-T. Wu. A probabilistic interpretation of ‘If—Then’. *The Quarterly Journal of Experimental Psychology*, 49(A):828–844, 1996.
- [15] D. Marr. *Vision. A computational investigation into the human representation and processing of visual information*. W. H. Freeman, San Francisco, 1982.
- [16] M. Oaksford and N. Chater. Conditional probability and the cognitive science of conditional reasoning. *Mind & Language*, 18(4):359–379, 2003.
- [17] M. Oaksford and N. Chater. *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press, Oxford, 2007.
- [18] K. Oberauer and O. Wilhelm. The meaning(s) of conditionals: Conditional probabilities, mental models and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29:680–693, 2003.
- [19] D. E. Over, C. Hadjichristidis, J. S. Evans, S. J. Handley, and S. A. Sloman. The probability of causal conditionals. *Cognitive Psychology*, 54:62–97, 2007.
- [20] N. Pfeifer and G. D. Kleiter. Nonmonotonicity and human probabilistic reasoning. In *Proceedings of the 6<sup>th</sup> Workshop on Uncertainty Processing*, pages 221–234, Hejnice, 2003. September 24–27<sup>th</sup>, 2003.
- [21] N. Pfeifer and G. D. Kleiter. Coherence and nonmonotonicity in human reasoning. *Synthese*, 146(1-2):93–109, 2005.
- [22] N. Pfeifer and G. D. Kleiter. Towards a mental probability logic. *Psychologica Belgica*, 45(1):71–99, 2005.
- [23] N. Pfeifer and G. D. Kleiter. Inference in conditional probability logic. *Kybernetika*, 42:391–404, 2006.
- [24] N. Pfeifer and G. D. Kleiter. Is human reasoning about nonmonotonic conditionals probabilistically coherent? In *Proceedings of the 7<sup>th</sup> Workshop on Uncertainty Processing*, pages 138–150, Mikulov, 2006. September 16–20<sup>th</sup>, 2006.
- [25] N. Pfeifer and G. D. Kleiter. Towards a probability logic based on statistical reasoning. In *Proceedings of the 11<sup>th</sup> IPMU Conference (Information Processing and Management of Uncertainty in Knowledge-Based Systems)*, pages 2308–2315, Paris, 2006. Edition E.D.K.
- [26] G. Politzer. Uncertainty and the suppression of inferences. *Thinking & Reasoning*, 11(1):5–33, 2005.
- [27] G. Politzer and G. Bourmaud. Deductive reasoning from uncertain conditionals. *British Journal of Psychology*, 93:345–381, 2002.
- [28] L. J. Rips. *The psychology of proof: Deductive reasoning in human thinking*. MIT Press, Cambridge, 1994.
- [29] G. Schurz. Non-monotonic reasoning from an evolution-theoretic perspective: Ontic, logical and cognitive foundations. *Synthese*, 1-2:37–51, 2005.
- [30] M. J. Smithson. *Human judgment and imprecise probabilities*. [www.sipta.org/documentation](http://www.sipta.org/documentation), 1997-2000.
- [31] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.