The Non-Use of Bayes Rule: Representative Evidence on Bounded Rationality

Thomas Dohmen

Armin Falk

David Huffman

Felix Marklein

Uwe Sunde

Sec09.004.pdf

Abstract

The Non-Use of Bayes Rule: Representative Evidence on Bounded Rationality*

The ability to process new information and to compute conditional probabilities is crucial for making appropriate decisions under uncertainty. In this paper, we investigate the capability of inferring conditional probabilities in a representative sample of the German population. Our results show that only a small fraction of the population responds consistently with Bayes' rule. Instead, most individuals either neglect the base probability, or the arrival of new information, in their responses. The probability to give normatively correct answers decreases with the level of education.

JEL classification: C90, D00, D10, D80, D81, D83, H00
Keywords: Bayes' Rule, Bayesian Updating, Base Rate Neglect, Bounded Rationality.

Thomas Dohmen
ROA
P.O. Box 616
6200 MD Maastricht
The Netherlands
t.dohmen@roa.unimaas.nl

David Huffman
Swarthmore College and IZA
Department 500 College Ave.
Swarthmore, PA 19081 USA
dhuffma1@swarthmore.edu

Uwe Sunde
University of St.Gallen, SEW-HSG
Varnbuelstr. 14
9000 St.Gallen, Switzerland
uwe.sunde@unisg.ch

Armin Falk
University of Bonn, IZA, and CEPR
Adenauerallee 24-42
53113 Bonn
Germany
armin.falk@uni-bonn.de

Felix Marklein
Bundesministerium der Finanzen
Wilhelmstrasse 97
10117 Berlin
Germany
felixmarklein@gmx.de

1 Introduction

Decision making under uncertainty often requires that agents combine new information with prior information they have. From a normative point of view, agents should employ the basic principles of statistical inference in this type of situation, and use all available information appropriately. In particular, they should apply Bayes' rule as a basis of their decision, in order to process the new information conditional on prior information. Experimental evidence suggests, however, that people often have difficulties with making judgments that involve conditional probabilities (Tversky and Kahneman 1971, Grether 1980, Charness and Levin 2005). The consequences of flawed probabilistic reasoning can be severe: for instance, an insufficient understanding of the importance of base rates has resulted in suicides of blood donors who were mistakenly tested positive for HIV in the 1980s, and incorrect use of probabilistic information has led to wrongful convictions in criminal cases (for detailed accounts, see Gigerenzer 2003). Despite an increased interest in behavioral deviations from the Bayesian updating model, insights about the determinants of biased probability judgment remain scarce. Moreover, little is known about the prevalence of judgment biases in the general population. Existing studies have been restricted to selected groups, such as university students, chemical workers, or medical doctors (Viscusi and O'Connor 1984, Hoffrage, Lindsey, Hertwig, and Gigerenzer 2000).

In this paper, we measure people's ability for Bayesian reasoning in a representative sample of the German population, consisting of more than 1,000 individual observations. This approach enables us to evaluate the prevalence of cognitive biases that are described in the experimental literature on a much broader sample of the population. The question we use to elicit respondents' ability in assessing conditional probabilities is framed as a judgment about weather conditions. This framing embeds the task in a context in which respondents are likely to be familiar with probabilities, as weather forecasts in TV and radio often refer to probabilistic measures. Participants received all information in form of natural frequencies, which has been shown to enhance participants' performance (Gigerenzer and Hoffrage 1995). The exact wording of the question was as follows (translated from German):

> *Imagine you are on vacation in an area where the weather is mostly sunny and you ask yourself how tomorrow's weather will be. Suppose that, in the area you are in, on average 90 out of 100 days are sunny, while it rains on 10 out of 100 days. The weather forecast for tomorrow predicts rain. On average, the weather forecast is correct on 80 out of 100 days. What do you think is the probability, in percent, that it is going to rain tomorrow?*[1]

---

[1] The original wording was: "Stellen Sie sich vor, Sie sind im Urlaub in einer Gegend, in der meistens schönes

Applying Bayes' rule, the solution is $\Pr(Rain|RainForecasted) = \Pr(R|RF) = \frac{\Pr(RF|R)\Pr(R)}{\Pr(RF)} = \frac{0.8*0.1}{0.9*0.2+0.1*0.8} = 0.308$.

Our results show that only a very small proportion of answers is close to the Bayesian solution. In particular, we find that *none* of the respondents gives the correct answer of 31%. A probability of 30% is indicated by 2.8% of the respondents, and the interval $31\% \pm 10\%$-points contains only 5.9% of all answers. The most frequent bias we find in the data is consistent with the so-called base rate neglect, a phenomenon that is well-established in the cognitive psychology literature (Bar-Hillel 1980, Tversky and Kahneman 1982, Koehler 1996). In terms of the question used here, this bias implies that respondents ignore the information about the base rate of rainy days and concentrate only on the probability with which the weather forecast is correct. The resulting estimate of 80% is given by 24.2% of the sample. Thus, in the setup under investigation, base rate neglect is over four times as frequent as answers that are located in the 10%-point interval around the Bayesian solution. Another bias that is well-known from the psychological literature is the tendency to answer with the pure base rate. In our setup, this would lead respondents to estimate the probability of rain tomorrow (given the weather forecast) to equal to the prior probability (i.e., the unconditional probability) of rain, ignoring the information provided by the weather forecast. This biased answer of 10% is given by 10.8% of respondents. Taken together, these results have two important implications: first, in a representative sample of the population, only a very small fraction of answers is close to the normatively correct answer. Second, deviations from the Bayesian solution appear to be systematic, as certain mistakes are by far more frequent than others.

In addition to the prevalence of different responses to cognitive biases, we analyze the determinants of biased probability judgment by relating respondents' probability estimates to individual background characteristics such as education, cognitive ability, gender, and age. In the economic literature, a common explanation for the existence of judgment biases is bounded rationality: people decide rationally under the constraint that they have cognitive limitations which may prevent them from finding the normatively correct solution (for an overview, see Conlisk 1996). Therefore, when analyzing the determinants of biased probability judgment, proxies for respondents' cognitive ability are of particular interest. A unique feature of our data set is that it includes two tests of cognitive ability that are based on sub-modules from one of the most widely used IQ tests, which were adapted to be implementable in a computer assisted

---

Wetter ist und Sie fragen sich, wie wohl morgen das Wetter sein wird. Angenommen an Ihrem Urlaubsort scheint im Durchschnitt an 90 von 100 Tagen die Sonne und an 10 von 100 Tagen regnet es. Für morgen lautet der Wetterbericht der Wettervorhersage dass es regnen wird. Die Wettervorhersage sagt das Wetter für den nächsten Tag im Durchschnitt an 80 von 100 Tagen richtig voraus. Was glauben Sie: Wie hoch ist die Wahrscheinlichkeit, ausgedrückt in Prozent, dass es morgen regnet?".

personal interview (CAPI). Moreover, the data contain information on respondents' years of schooling and on whether they hold a university degree. As control variables, respondents' age and gender are included.

Our empirical analysis of the determinants of biased probability judgment leads to surprising results: a high level of education does not increase the likelihood of giving an answer that is close to the Bayesian solution. Rather, we find that highly educated respondents are *more* likely to exhibit base rate neglect. These effects are particularly pronounced for people with an academic background: among respondents who hold a university degree, 37.9% give the base rate neglect estimate of 80%, and 2.6% give an answer that is in the 10%-point interval around the Bayesian solution. In contrast, among respondents without a university degree, 22.4% are prone to base rate neglect, and 6.3% give an answer in the Bayesian interval. Stratifying the sample by years of schooling leads to very similar results, with the share of base rate neglect being much higher among respondents with more than 10 years of schooling (31.5% vs. 20.6%). We confirm the statistical significance of our findings in a number of regressions, controlling for respondents' background characteristics. In particular, marginal effect estimates from probit regressions reveal that the probability to commit base rate neglect is roughly 15%-points higher for respondents who hold a university degree, controlling for cognitive ability, age, and gender. In a similar vein, one additional year of schooling leads to a 2%-point increase in the probability to commit base rate neglect. Furthermore, our estimates show that the impact of cognitive ability is very similar to the impact of education: a higher score in the cognitive ability measures significantly increases the probability to commit base rate neglect and leads to estimates that are further away from the normative solution. A series of robustness checks shows that our results continue to hold when estimating multinomial logit models and when using a linear measure of respondents' biases.

From a methodological viewpoint, the design of our study allows us to draw conclusions about the general population that cannot always be obtained with non-representative subject pools. Unfortunately, an adequate comparison of our findings to results from previous research is difficult, as most psychological studies do not report the full distribution of answers. Rather, their focus is either on the share of base rate neglect in the data, or on the share of Bayesian answers, depending on the respective research question. We discuss the results of our study in light of the existing literature in section 3.4.

The remainder of this paper is organized as follows: Section 2 contains a description of the data, and Section 3 presents the results. Section 4 concludes.

2   Data

The data set under investigation is a representative sample of the population living in Germany aged 16 years and older. The data were collected in June and July 2005 through the professional interview group TNS Infratest. Households were contacted by interviewers according to the Random Route Method (see Fowler 2002) and one person per household was surveyed. The interview procedure is identical to that used by the widely used German Socio-Economic Panel Study (SOEP). Respondents were interviewed in their homes. All interviewers used the CAPI procedure (Computer Assisted Personal Interview), administering questions and collecting answers with the help of a notebook computer.

Respondents' ability to infer conditional probabilities is elicited using a question about weather conditions mentioned in the introduction. Respondents had to answer with a number between 0% and 100%. Alternatively, they could give the answer "I don't know". Although the problem may seem complicated at first, an easy way to solve it is by finding the correct information partition. Out of 100 days, the sun shines on 90 days and it rains on 10 days. For the 90 sunny days, the weather forecast predicts sunshine on 72 days and rain on 18 days. For the 10 rainy days, it predicts sunshine on 2 days and rain on 8 days. Thus, given that the forecast predicts rain, the probability that it will actually rain is $8/(18 + 8) \approx 31\%$. The normative solution to this question is therefore a rain probability of about 31%. The structure of this question is very similar to the "cab problem", introduced by Kahneman and Tversky (1972). However, in order to identify correct answers without ambiguity, we modified the base rate in our task such that the Bayesian solution is neither close to the base rate (10%), nor to the middle of the answer space (50%), which might be a focal point.[2] In total, 988 participants answered the probability question.

The main goal of this paper is an assessment of the prevalence of probability judgment biases and an analysis of their determinants. We deliberately do not address the question of whether a certain bias leads to "good" or "bad" estimates, as this depends crucially on the specific parameters of the decision environment. For instance, when looking at the bias of base rate neglect, it follows directly from Bayes' rule that the difference between the Bayesian posterior and the base rate neglect estimate becomes very large for base rates close to 0%, whereas it becomes negligible for base rates close to 50%. To illustrate this point, assume the base rate of rainy days in our weather task was not 10%, but 5%. The resulting posterior would be 17.4%, which is far away from the estimate of 80% that is obtained by base rate neglect. In contrast,

---

[2]In the cab problem, the probability with which the additional information is correct is 80% (as it is in our task), but the base rate is 15% instead of 10%, leading to a Bayesian posterior of 41%. This can make interpretation of the results problematic, as the correct solution is then relatively close to the focal point of 50%.

a base rate of 45% would lead to a Bayesian posterior of 76.6%, which is very close to the base rate neglect estimate. Thus, from a decision-making perspective, base rate neglect can be a useful heuristic in environments where the base rate of an event is close to 50%. The question of whether respondents realize under which circumstances a certain heuristic is more useful than another is beyond the scope of this paper.

In addition to the elicitation of conditional probability judgments, the data contain a number of background characteristics. A novel feature of the survey is that it elicited measures of respondents' cognitive ability. One measure elicits respondents' perceptual speed through a test in which participants have to find correspondences between symbols and digits. The other measure is a word fluency test that provides a proxy of respondents' general knowledge. The design of these measures was based on the two-component theory of cognitive ability, originating from research in developmental psychology (Lang 2005). According to this theory, cognitive ability can be broadly divided into cognitive mechanics and cognitive pragmatics (Baltes, Lindenberger, and Staudinger 2006, Lindenberger and Baltes 1997). The mechanics of cognition (fluid intelligence) reflect fundamental organizational properties of the central nervous system (Singer 1995). In contrast, the cognitive pragmatics (crystallized intelligence) reflect the knowledge-based domain of cognitive ability. Examples for the mechanics of cognition are the speed, the accuracy, and the coordination of cognitive processing operations, whereas examples for the pragmatics are reading and writing skills, educational qualifications, and professional skills. Respondents' performance in the domain of cognitive mechanics was assessed via a digit-symbol test that has been designed to measure perceptual speed. For this test, respondents had to match the correct digit to symbols on the computer screen. They had to match as many digit-symbol pairs as possible within a time frame of 90 seconds. The CAPI method allowed us to measure performance directly, registering decisions through a software that was running in the background. To measure performance in the area of cognitive pragmatics, a word fluency test was developed. In this test, participants were asked to name as many distinct animals as possible in 90 seconds of time.[3]

Sample statistics are shown in Table 1. About 53.7% of the sample are female, average age is 47.6 years, and respondents have on average 10.3 years of schooling. The measures for cognitive ability have a mean value of 28.3 (perceptual speed test) and 22.8 (word fluency test).

---

[3]For the digit-symbol test, the CAPI procedure proved very effective, resulting in an error rate of virtually zero. In the word fluency test, the error rate was slightly higher, due to the fact that the design demanded a fairly high level of concentration from the interviewers, who had to recall which animals had already been mentioned and which had not. To correct imprecisions in the data, we evaluated the interviewers' comments of all tests that were ended early. Whenever the comments hinted at a discrepancy between the number of animals named by the respondent and the number recorded by the computer, we replaced the recorded number by the interviewer's estimate of the actual number. Results from a re-test with audio recordings of respondents' answers revealed a fairly high accuracy of the interviewers' estimates (see also Dohmen, Falk, Huffman, and Sunde 2007).

Table 1: Sample Statistics.

Variable	average	std. dev.	num. obs.
1 if female	0.537	(0.499)	1,012
Age	47.555	(18.371)	1,011
Word fluency test (raw data)	22.807	(10.986)	853
Word fluency test (standardized)	0	(1)	853
Perceptual speed test (raw data)	28.309	(9.890)	903
Perceptual speed test (standardized)	0	(1)	903
Years of schooling	10.328	(1.829)	965
1 if university degree	0.118	(0.323)	1,009
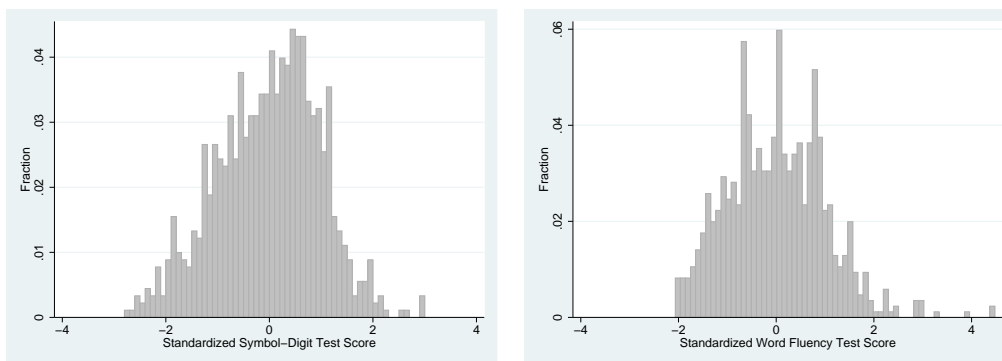
(a) Histogram of standardized score in perceptual speed test. Respondents had to match as many digits and symbols as possible in 90 seconds of time.

(b) Histogram of standardized score in word fluency test intelligence. Respondents had to name as many distinct animals as possible in 90 seconds of time.

Figure 1: Measures of Cognitive Ability.

For the empirical analysis, both cognitive ability measures are standardized to have mean zero and standard deviation one. Figure 1 presents histograms of the two standardized measures. We also construct a combined measure of cognitive ability by averaging the two standardized measures and then standardizing the resulting average. Combining the measures is standard procedure, and should help reduce measurement error.

3 Results

3.1 Prevalence of Biased Probability Judgment

A first look at the results shows that respondents' estimates regarding the probability of rain are spread out over the full range of the answer space (Figure 2). Still, we find that none of the 988

respondents gives the Bayesian answer of 31%. A rain probability of 30% is indicated by 2.8% of the respondents, and only 3.0% of answers are in the interval 31% ± 5%-points. Even the interval 31% ± 10%-points contains only a small share (5.9%) of all answers. Taken together, these findings show that a large majority of the respondents has difficulties with making Bayesian judgments in the setting at hand, as their estimates are far away from the normatively correct solution.
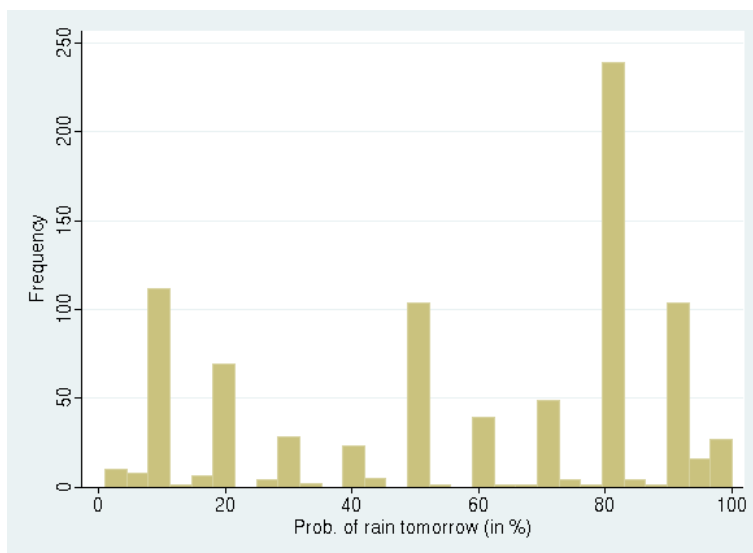

Figure 2:
Notes: Answers to Conditional Probability Question. The normatively correct solution is a rain probability of about 31%. The 129 observations for the answer "I don't know" are not shown.

Next, we analyze the answers in the sample that deviate from the normative solution. Table 2 provides an overview of typical judgment biases and shows their prevalence in our data. Recall that a common bias described in the literature on conditional probability judgment is base rate neglect. We find that, in our setting, the corresponding estimate of 80% is given by 24.2% of the respondents. Thus, almost a quarter of the respondents chooses in line with the most prominent prediction for biased probability judgment. A second bias that is often reported in the literature is respondents' tendency to focus only on the base rate, thereby ignoring the information provided in addition to the base rate probability. The corresponding estimate of 10% is given by 10.8% of the sample, and the 10%-point interval around the pure base rate estimate contains 20.7% of all answers.

In sum, we find that in our setup the share of respondents who exhibit base rate neglect is about four times as large as the share of respondents whose answer is located in the Bayesian interval. Moreover, the pure base rate interval contains more than three times as many answers as the Bayesian interval. These results are a clear indication that the pattern of incorrect answers

Table 2: Bayesian Solution vs. Typical Judgment Biases.

	Solution process	Resulting answer	% of answers in sample
Bayesian:	$\frac{\Pr(RF|R)*\Pr(R)}{\Pr(RF)} = \frac{0.8*0.1}{0.9*0.2+0.1*0.8} = 0.308$	30% (20%,40%]	2.8% 5.9%
Pure base rate:	$\Pr(R) = 0.1$	10% [0%,20%]	10.8% 20.7%
Base rate neglect:	$\Pr(RF|R) = 0.8$	80%	24.2%
Other:			49.2%

Notes: Category "Other" consists of all answers in the intervals (40%,80%), (80%, 100%], and "I don't know". "R" denotes "Rain tomorrow" while "RF" denotes "Rain forecasted"

is non-random. It is very unlikely that the observed deviations from the normative solution are due to an error source like lack of concentration or lack of motivation. Rather, the data pattern indicates that respondents' answers are driven by systematic biases in probabilistic judgment.

3.2 Determinants of Biased Probability Judgment

If probability judgment biases occur due to boundedly rational behavior, we should expect a significant relationship between respondents' probability estimates and proxies for their cognitive ability. We therefore proceed to investigate the interplay between respondents' answers on the one hand and their performance in the cognitive ability tests and their education levels on the other. The descriptive statistics in Table 3 reveal a striking finding: stratifying the sample according to years of schooling indicates that highly educated people are much more likely to commit base rate neglect, whereas they do not have a higher likelihood to give an answer that is close to the Bayesian solution.

We find that 31.5% of those with more than 10 years of schooling exhibit base rate neglect, and 4.9% give answers in the interval that contains the Bayesian solution. In contrast, among people with at most 10 years of schooling, only 20.6% commit base rate neglect, and 6.3% give an answer in the Bayesian interval. The effect of education is even more pronounced if we focus our analysis on respondents with a university education: among those who hold a university degree, 37.9% are prone to base rate neglect, and 2.6% give an answer in the Bayesian interval. In contrast, among people without university degree, 22.4% exhibit base rate neglect, and 6.3%

Table 3: Probability Judgement by Sub-groups.

	Bayesian	Pure base rate	Base rate neglect	Other	Total
All	58 (5.9%)	205 (20.7%)	239 (24.2%)	486 (49.2%)	988 (100%)
Years of schooling > 10	16 (4.9%)	68 (21.0%)	102 (31.5%)	138 (42.6%)	324 (100%)
Years of schooling ≤ 10	42 (6.3%)	137 (20.6%)	137 (20.6%)	348 (52.4%)	664 (100%)
University degree	3 (2.6%)	30 (25.9%)	44 (37.9%)	39 (33.6%)	116 (100%)
No univ. degree	55 (6.3%)	175 (20.1%)	195 (22.4%)	445 (51.2%)	870 (100%)
Word fluency > 0	26 (4.6%)	114 (20.3%)	151 (26.9%)	270 (48.1%)	561 (100%)
Word fluency ≤ 0	32 (7.5%)	91 (21.3%)	88 (20.6%)	216 (50.6%)	427 (100%)
Perceptual speed > 0	24 (4.3%)	112 (19.9%)	147 (26.1%)	281 (49.8%)	564 (100%)
Perceptual speed ≤ 0	34 (8.0%)	93 (21.9%)	92 (21.7%)	205 (48.4%)	424 (100%)
Average cognitive ability score > 0	28 (4.7%)	121 (20.3%)	155 (26.0%)	292 (49.0%)	596 (100%)
Average cognitive ability score ≤ 0	30 (7.7%)	84 (21.4%)	84 (21.4%)	194 (49.5%)	392 (100%)
Female	31 (5.8%)	117 (22.0%)	116 (21.8%)	268 (50.4%)	532 (100%)
Male	27 (5.9%)	88 (19.3%)	123 (27.0%)	218 (47.8%)	456 (100%)
Age < 50	27 (4.9%)	107 (19.3%)	158 (28.5%)	262 (47.3%)	554 (100%)
Age ≥ 50	31 (7.1%)	98 (22.6%)	81 (18.7%)	224 (51.6%)	434 (100%)

Notes: Table shows number of answers in each answer category. "Bayesian" are answers in the interval (20%, 40%] around the Bayesian solution of 30.8%. "Pure base rate" are answers in the interval [0%, 20%] around the pure base rate of 10%. "Base rate neglect" are answers equal to 80%. Category "Other" consists of all answers in the intervals (40%, 80%), (80%, 100%], and "I don't know". Cognitive ability measures are standardized.

give an answer in the Bayesian interval. Very similar results obtain when we split the sample at the median according to the cognitive ability measures.[4] A look at the control variables gender and age reveals that there is a gender effect, with men being more likely to commit base rate neglect (27.0% vs. 21.8%). A similar pattern can be seen for age, with younger people being more likely to neglect the base rate (28.5% vs. 18.7%).

In the following, we use regression analysis in order to control for respondents' background characteristics and to assess whether our findings are statistically significant. In a first step, we analyze the determinants of base rate neglect, as this is both the most commonly described phenomenon in the literature and the most frequent bias in our sample. We estimate a probit model with the dependent variable being equal to one if a respondent makes the mistake of base rate neglect. As explanatory variables we include the measures of cognitive ability, age, gender, and years of schooling. The results are presented in Table 4. We find that the signif-

---

[4]In particular, in the group with high cognitive ability, the share of people who give an answer in the Bayesian interval is about 4%, whereas it is about 8% in the group of people with lower cognitive ability.

icant determinants of whether a respondent is prone to base rate neglect are cognitive ability and education. Remarkably, all these measures have positive coefficients, meaning that people with better cognitive ability and with more education are more likely to neglect the base rate information. For instance, the estimates in columns (3) and (4) of Table 4 show that an increase in the average cognitive ability measure by one standard deviation leads to an increase in the probability to commit base rate neglect by about 5.6 percentage points ($p < 0.01$), whereas an additional year of schooling leads to an increase in this probability by another 2%-points ($p < 0.1$). In contrast, in most specifications the effect of age and gender is not significant once we control for cognitive ability and education.

Table 4: Determinants of Base Rate Neglect

Dependent variable: =1 if answer is 80% (base rate neglect), =0 otherwise

	(1)	(2)	(3)	(4)	(5)
Word fluency score	0.040** [0.016]				
Speed recognition score		0.035* [0.018]			
Average cognitive ability score			0.056*** [0.018]	0.056*** [0.017]	0.054*** [0.018]
1 if female	-0.059* [0.031]	-0.050 [0.030]	-0.054* [0.032]	-0.049 [0.031]	-0.053* [0.032]
Age	-0.001 [0.001]	-0.001 [0.001]	0.000 [0.001]	-0.000 [0.001]	-0.000 [0.001]
Years of Schooling	0.019** [0.009]	0.023*** [0.008]	0.017* [0.009]		0.006 [0.011]
1 if university degree				0.138*** [0.053]	0.116* [0.062]
Observations	806	851	768	809	768
Prob > χ^2	0.0001	0.0002	0.0001	0.0000	0.0001
Pseudo R^2	0.025	0.023	0.026	0.029	0.031

Notes: Probit estimates, marginal effects evaluated at the mean with standard errors in brackets. Dependent variable is a dummy equal to one if respondent's answer is 80% (base rate neglect). Cognitive ability measures are standardized to have mean 0 and standard deviation 1. Significance at the 1%-, 5%-, and 10%-level is denoted by ***, **, and *, respectively.

The descriptive part of the analysis has already stressed that base rate neglect is especially widespread among respondents who hold a university degree. This is an important finding, as previous research on probability judgment has relied heavily on subject pools consisting of university students and of academically trained professionals (e.g., in medicine). In contrast, the representative nature of our study entails that only 12% of the respondents have an academic

background. The data set therefore allows us to assess whether the behavior of people with a university education can be generalized to the behavior of the population at large. To study this question in a regression framework, we define a dummy variable that is equal to one if the respondent holds a university degree, and equal to zero otherwise. Results from probit regressions with the base rate neglect dummy as dependent variable are shown in Columns (4) and (5) of Table 4. The results are striking: controlling for gender, age, and cognitive ability, people with a university degree are about 14 percentage points more likely to commit base rate neglect. Given that in the total sample the share of people who commit base rate neglect is about 24%, this is a substantial effect that is difficult to be reconciled with existing theories of human decision making.

We have seen that, next to base rate neglect, a common judgment bias is respondents' tendency to answer with the pure base rate. In our sample, 20.7% of all answers are in the interval [0%,20%] where the pure base rate of 10% is located. Still, it is not clear what the determinants of this bias are. As the descriptive statistics in Table 3 have shown, stratifying the sample according to cognitive ability, age, gender, and education results only in very little variation in the share of respondents whose estimate is located in the pure base rate interval. For instance, this share is 21.0% among respondents with more than 10 years of schooling, whereas it is 20.6% among respondents with 10 years of schooling or less. Similarly, for all other measures the share of respondents in the pure base rate interval is always relatively close to 20%, indicating that these measures cannot explain what the determinants of giving an estimate close to the pure base rate are. Analogous to the previous analysis for base rate neglect, we performed probit regressions where the dependent variable is a dummy equal to one for answers in the pure base rate interval and equal to zero otherwise. In these regressions, none of the explanatory variables has a significant impact. The same is true for probit regressions with the dependent variable being equal to one if a respondent's answer is exactly 10%.

3.3 Robustness Checks

In this subsection we perform two sets of robustness checks. First, in order to address the interplay of respondents' answers and their background characteristics in more detail, we estimate multinomial logit models with the same four answer categories we investigated earlier in the descriptive analysis.[5] These regressions allow us to test whether differences between the answer categories are statistically significant. Second, we estimate regressions in which the dependent variable is the linear distance between a respondent's answer and the Bayesian solution. Using

---

[5]Results do not depend on the specific categorization we choose. Multinomial logit regressions with a more detailed answer categories are available upon request.

Table 5: Robustness Check - Multinomial Analysis

Dependent variable: =1 if "Other", =2 if BRN, =3 if PBR, =4 if Bayesian

	(1)			(2)		
	Base rate neglect	Pure base rate	Bayesian	Base rate neglect	Pure base rate	Bayesian
Average cognitive ability score average iq	0.282***	-0.003	-0.136	0.264**	-0.016	-0.152
	[0.102]	[0.110]	[0.198]	[0.103]	[0.110]	[0.199]
Years of Schooling	0.102**	0.081	-0.261**	0.020	0.018	-0.328***
	[0.050]	[0.054]	[0.104]	[0.060]	[0.064]	[0.116]
1 if university degree				0.843***	0.670*	0.925
				[0.315]	[0.344]	[0.728]
1 if female	-0.227	0.199	0.020	-0.225	0.202	0.036
	[0.178]	[0.191]	[0.340]	[0.179]	[0.192]	[0.341]
Age	0.003	0.007	0.002	-0.001	0.004	-0.000
	[0.006]	[0.006]	[0.010]	[0.006]	[0.006]	[0.011]
Constant	-1.683***	-2.070***	0.239	-0.756	-1.377*	0.924
	[0.627]	[0.673]	[1.159]	[0.721]	[0.772]	[1.256]
N. Obs.		768			768	
Prob > χ^2		0.0001			0.0000	
Pseudo R^2		0.021			0.026	

Notes: Multinomial logit estimates, standard errors in parentheses. Dependent variable =1 if "Other" (answer either in the intervals (40%, 80%), (80%, 100%], or "Don't know"), =2 if Base rate neglect (answer equal to 80%), =3 if Pure base rate (answer in [0%,20%]), and =4 if Bayesian (answer in (20%, 40%]). The reference category is "Other". Cognitive ability measures are standardized to have mean 0 and standard deviation 1. Significance at the 1%, 5%, and 10% level is denoted by ***, ** and *, respectively.

this continuous measure enables us to analyze the determinants of biased probability judgment by taking into account whether deviations are either closer or further away from the correct solution. This might be relevant, e.g., in settings where estimates that are further away from the correct solution entail a higher cost for the decision maker.

In the multinomial regression analysis, the dependent variable indicates in which of the four categories a respondents' answer is located: Bayesian, pure base rate, base rate neglect, or "other". The base category in the regressions is "other". This allows us to address the factors that determine whether a respondent gives a Bayesian estimate or whether he displays a typical judgment bias. The estimation results support all the earlier findings from the descriptive analysis (see Table 5). For the specification in panel (1), the coefficient for years of schooling indicates that respondents with more education have a significantly higher probability of base

rate neglect ($p < 0.05$) and a significantly lower probability of giving an answer that is located in the Bayesian interval ($p < 0.05$). Note that these results were obtained from a regression which included the cognitive ability measures that combine the information from the two tests. The results also hold when we include the perceptual speed measure or the verbal fluency measure as regressors.[6] Higher cognitive ability makes it more likely to neglect the base rate. Conditional on education and cognitive ability, the estimates for age and gender are not significant.

Next, we address our earlier finding that people with an academic background have a strong tendency to commit base rate neglect. The results of multinomial logit regressions in panel (2) of Table 5, which include a dummy equal to one if the respondent holds a university degree, confirm our earlier finding that people with an academic background have a strong tendency to neglect the base rate ($p < 0.01$) and are more likely to answer with the pure base rate ($p < 0.1$).

Table 6: Robustness Check - Linear Distance.

Dependent variable: distance from 30.8% (in %-points)

	(1)	(2)	(3)	(4)
Word fluency score	1.412*			
	[0.721]			
Speed recognition score		1.731**		
		[0.798]		
Average cognitive ability score			1.766**	1.791**
			[0.802]	[0.805]
1 if female	-0.502	-0.203	-0.560	-0.562
	[1.385]	[1.338]	[1.402]	[1.403]
Age	-0.109***	-0.091*	-0.090*	-0.085*
	[0.042]	[0.046]	[0.046]	[0.047]
Years of Schooling	0.810**	0.918**	0.765*	0.860*
	[0.383]	[0.371]	[0.391]	[0.460]
1 if university degree				-0.964
				[2.455]
Constant	32.592***	30.504***	32.176***	31.105***
	[4.791]	[4.649]	[4.827]	[5.546]
N. Obs.	716	766	692	692
R-squared	0.030	0.038	0.034	0.034

Notes: OLS estimates, standard errors in brackets. Dependent variable is the linear distance (in %-points) between a respondent's answer and the Bayesian solution of 30.8%. By definition, all observations where a respondent answered "I don't know" are excluded from the following analysis. Cognitive ability measures are standardized to have mean 0 and standard deviation 1. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively. Respondents who answered "I don't know" are not included.

In a final step, we present results from a series of OLS regressions where the dependent

---

[6]Results are available upon request.

variable is the linear distance from the normative solution (in %-points). This approach allows us to take the cost of an error into account. In particular, one might assume that estimates that are further away from the correct solution entail a higher cost to the decision maker. Estimation results are shown in Table 6. We find that both the cognitive ability measure and years of schooling have positive coefficients and are highly significant in all specifications. This implies that, in the setup under investigation, higher cognitive ability and more education lead respondents to estimates that are further away from the Bayesian solution. For instance, the estimates in columns (3) and (4) indicate that a one standard deviation increase in the cognitive ability measure translates into an increase in the respondent's bias by 1.8 percentage points ($p < 0.05$), whereas an additional year of schooling increases this bias by another about 0.8-0.9 percentage points ($p < 0.1$). Notably, the measure for perceptual speed has a stronger detrimental effect than the word fluency measure. Among the control variables, age has a significantly negative impact in most specifications, indicating that, on average, older people tend to be further away from the Bayesian solution.

3.4 Discussion

Our results complement the existing literature in several dimensions and shed new light on the prevalence of agents' failure to deal with conditional probabilities. In a seminal study by Casscells, Schoenberger, and Graboys (1978), 60 students and staff from Harvard Medical School were asked to estimate the probability that a patient actually has a disease, given a positive result in a diagnostic test. Parameters of the question were chosen such that the normatively correct answer is approximately 2%, whereas base rate neglect leads to an estimate of 95%. Results show that about 18% of the participants give the correct answer. In contrast, almost half of the respondents exhibit base rate neglect. The average estimate in the sample is 56%. Eddy (1982) reports from a similar study that among 100 physicians, 95 exhibit base rate neglect. In an overview article, Tversky and Kahneman (1982) state that base rate neglect is typically both the modal answer and the median answer in Bayesian updating tasks. Unfortunately, their article neither contains figures regarding the exact share of respondents committing base rate neglect, nor does it report the typical share of normatively correct answers. Gigerenzer and Hoffrage (1995) investigate 15 distinct Bayesian updating tasks in order to show that presenting information in natural frequencies instead of probabilities enhances respondents' tendency to apply Bayesian reasoning. The experiments were conducted with 60 undergraduates (mainly psychology majors) from the University of Salzburg, Austria. In the updating task that is most similar to the one we use in this paper, the authors find that about 15% of the sample give the Bayesian answer. However, no information is reported about the distribution of non-normative

14

answers. In particular, nothing is said about the share of subjects who exhibit base rate neglect or who answer with the pure base rate. A study that reports a full picture of the answers in a Bayesian updating task is provided by Cosmides and Tooby (1996). The task used in their study is a standard medical diagnosis problem, where the base rate of the disease is 0.1% and the false positive rate of the diagnostic test is 5%. The Bayesian posterior for having the disease, conditional on a positive test result, is approximately 2%. In their sample of 25 Stanford undergraduates, base rate neglect is exhibited by 14 subjects (56% of the sample), whereas 3 subjects (12%) give the Bayesian answer, and another 3 subjects (12%) answer with the pure base rate. Closest to our analysis is a study by Stanovich and West (1998) who analyze the relationship between respondents' cognitive ability and their assessment of probabilities in a sample of 211 psychology undergraduates.

In comparison to this previous research, the share of base rate neglect in our sample is much lower than it is in studies with non-representative subject pools. Taking our analysis of the determinants of probabilistic judgment biases at face value, one potential reason for this difference are the different education levels of participants in our study and participants in previous research. The existing studies have mostly been conducted with academically trained individuals. In contrast, only 12% of the respondents in our representative data set hold a university degree. Finding a much higher share of base rate neglect in samples that consist entirely of academically trained subjects is in line with our empirical findings concerning the determinants of biased probability judgment, as our results have shown that the likelihood of being prone to this particular bias rises significantly with respondents' education level. These findings are important: one might be tempted to conclude from previous studies that base rate neglect must be an extremely widespread phenomenon, since tests with highly sophisticated students and professionals should constitute a lower bound for the prevalence of flawed decision making in the general population. Our results are not consistent with this view. In light of our findings, it is likely that previous studies with highly educated subjects resulted in an *upper* bound for the prevalence of base rate neglect.

When considering the share of Bayesian answers and the share of answers that give the pure base rate, generalizing the results from student subject pools to the population at large seems more appropriate. In these two domains, differences between our findings and the existing literature are relatively small. This is again in line with our empirical analysis of the underlying determinants, as we find that respondents' level of education has neither a significant impact on their likelihood to give a Bayesian estimate, nor on their likelihood of answering with the pure base rate.

By investigating not only the role of education, but of direct measures of cognitive ability, for

processing conditional probabilities, our paper complements work by Stanovich and West (1998). In their study, they analyze a sample of 211 psychology undergraduates from a medium-sized U.S. university and investigate the relationship between respondents' cognitive ability and their assessment of probabilities. Subjects had to solve two conditional probability questions: the cab problem and a standard medical diagnosis task. As a proxy for cognitive ability, respondents were asked to report their test scores in the Scholastic Aptitude Test (SAT), a standardized test for college admissions in the United States.[7] The parameters of the cab problem are chosen such that the base rate is 15% and the additional info is correct with 80%, leading to a Bayesian posterior of 41%. In contrast to our study, subjects were not asked for a precise estimate but had to choose their answer from six alternative intervals: less than 10%, 10-30%, 30-50%, 50-70%, 70-90%, more than 90%. To report the findings, the authors construct three categories: [0%,30%) for answers that rely too heavily on the base rate (n=38), [30%,70%] for Bayesian answers (n=45) and (70%,100%] for answers that tend to neglect the base rate (n=128, see Table 4 of their paper).[8] The results show that the average SAT scores for the three answer categories are rather similar: 1096, 1129, and 1145, respectively. None of the SAT score differences is significant at conventional levels. In contrast, outcomes of the medical diagnosis task lead to significant results. Here, subjects who score higher on measures of cognitive ability are more likely to exhibit base rate neglect, which seems at first sight to be in line with the findings of our study. Unfortunately, the medical diagnosis task from which this claim is derived does not allow for a distinction between Bayesian answers and answers that are equal to the pure base rate: the parameters are chosen such that the base rate is 0.1% and the additional information is correct with 95%. The resulting Bayesian posterior is roughly 2%. Again, respondents could choose their answer from the same six intervals used in the cab problem. For the data analysis, the authors construct three categories: answers in [0%,10%) are interpreted as Bayesian estimates (n=36), answers in [10%,90%] are defined as "intermediate" (n=57), and answers in (90%,100%] are said to neglect the base rate (n=118, see Table 5 of their paper). The results show that average SAT scores for the three categories are 1103, 1109, and 1153, indicating that test scores of subjects in the Bayesian interval and in the intermediate interval are very similar, whereas test scores in the base rate neglect interval are significantly higher ($p < .025$). A potential problem of this analysis is that Bayesian answers ("2%") and answers that give only the pure

---

[7]Additional proxies for cognitive ability were elicited through specifically designed problems from the psychology literature, e.g., the so-called Raven Matrices and the Nelson-Denny comprehension measure. For simplicity, we focus on the results that relate to the SAT test score measure. Results for the other measures of cognitive ability were very similar.

[8]Unfortunately, the intervals used in the data analysis are relatively large, and answers are not reported in more detail. For instance, for an appropriate comparison of their results to our findings it would have been interesting to know the share of subjects in the interval 30-50%, as these are closest to the Bayesian solution of 41%.

base rate ("0.1%") are pooled together in the same category. Thus, associating the low average SAT score for this group exclusively with respondents who give a Bayesian estimate might be misleading. It could as well be that average SAT scores are lower for people who answered with the pure base rate. In fact, in the cab problem mentioned above, the average SAT score is indeed lowest in the pure base rate category. By using a modified question that allows us to separate the different response categories, and using a representative subject pool, our study allows us to shed new light on these findings.

4 Conclusion

In this paper, we have investigated people's ability for probabilistic judgment in a representative sample of the German population. In particular, we have used a task that has been specifically designed to test whether respondents are able to form conditional probability estimates. Our results have shown that only very few answers are in the vicinity of the Bayesian solution. Deviations from the normatively correct answer appear to be systematic and the two most frequently described biases in the literature, base rate neglect and pure base rate, feature prominently in the data. The analysis of the determinants of biased probability judgment has led to surprising results: highly educated respondents do not have a higher likelihood of giving a correct answer. Rather, they are significantly more likely to commit the mistake of base rate neglect. In a similar vein, measures for cognitive ability have a positive and highly significant impact on the probability that people commit base rate neglect.

An important implication of our findings is that the existing research on decision making with conditional probabilities does not translate one-to-one to the general population, as previous studies have mostly been conducted with university students and with academic professionals. It has to be noted, however, that our results may not be independent of the task structure. In our task, base rate neglect is a bias that leads to estimates that are far away from the Bayesian solution. In contrast, in situations where the base rate neglect outcome is closer to the normatively correct solution, it might be a useful heuristic to neglect the base rate. Taking our results at face value one would therefore expect that highly educated respondents perform better than other respondents in settings where the base rate neglect estimate is close to the Bayesian solution.

Our results appear surprising in light of standard theories of human decision making, as all our measures for education and for cognitive ability have a significant detrimental effect on respondents' probability judgment. Even approaches that allow for bounded rationality, e.g., by incorporating deliberation cost (Smith and Walker 1993), appear problematic in light of our

findings. In a cognitive task as complex as the one we use in this paper, one would expect deliberation cost to be relatively high for people with less formal education. In contrast, for highly educated people deliberation cost should be relatively low. Other things equal, this reasoning would imply that more educated people perform better in assessing conditional probabilities. Our results indicate the contrary, as education, in particular university education, increases the likelihood that respondents are led astray in the probability judgment task. An identification of the exact channels which are responsible for the detrimental effect of education and cognitive ability on Bayesian judgment constitutes a fascinating area for future research.

References

BALTES, P. B., U. LINDENBERGER, AND U. M. STAUDINGER (2006): "Life span theory in developmental psychology," in *Handbook of child psychology: Vol. 1. Theoretical models of human development (6th ed.)*, ed. by W. Damon, and R. M. Lerner, pp. 569–664. Wiley, New York.

BAR-HILLEL, M. (1980): "The Base-Rate Fallacy in Probability Judgments," *Acta Psychologica*, 44(3), 211–233.

CASSCELLS, W., A. SCHOENBERGER, AND T. GRABOYS (1978): "Interpretation by Physicians of Clinical Laboratory Results.," *New England Journal of Medicine*, 299(18), 999–1001.

CHARNESS, G., AND D. LEVIN (2005): "When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect," *American Economic Review*, 95, 1300–1309.

CONLISK, J. (1996): "Why Bounded Rationality?," *Journal of Economic Literature*, 34, 669–700.

COSMIDES, L., AND J. TOOBY (1996): "Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions From the Literature on Judgment Under Uncertainty," *Cognition*, 58(1), 1–73.

DOHMEN, T., A. FALK, D. HUFFMAN, AND U. SUNDE (2007): "Are Risk Aversion and Impatience Related to Cognitive Ability?," *IZA Discussion Paper No. 2735*.

EDDY, D. M. (1982): "Probabilistic Reasoning in Clinical Medicine: Problems and Opportunities," in *Judgment Under Uncertainty: Heuristics and Biases*, ed. by D. Kahneman, P. Slovic, and A. Tversky, pp. 249–267. Cambridge University Press, Cambridge.

FOWLER, F. J. (2002): *Survey Research Methods*. Sage Publications, London.

GIGERENZER, G. (2003): *Calculated Risks: How to Know When Numbers Deceive You*. Simon & Schuster, New York.

GIGERENZER, G., AND U. HOFFRAGE (1995): "How to Improve Bayesian Reasoning without Instructions: Frequency Formats," *Psychological Review*, 102(2), 684–704.

GRETHER, D. M. (1980): "Bayes Rule as a Descriptive Model: The Representativeness Heuristic," *Quarterly Journal of Economics*, 95(3), 537–557.

HOFFRAGE, U., S. LINDSEY, R. HERTWIG, AND G. GIGERENZER (2000): "Communicating Statistical Information," *Science*, 290(5500), 2261–2262.

KAHNEMAN, D., AND A. TVERSKY (1972): "On Prediction and Judgment," *ORI Research monograph*, 12.

KOEHLER, J. J. (1996): "The Base Rate Fallacy Reconsidered: Descriptive, Normative and Methodological Challenges," *Behavioral and Brain Sciences*, 19(1), 1–53.

LANG, F. R. (2005): "Erfassung des kognitiven Leistungspotenzials und der 'Big Five' mit Computer-Assisted-Personal-Interviewing (CAPI): Zur Reliabilitt und Validitt zweier ultrakurzer Tests und des BFI-S," *DIW Berlin, Research Notes, 9/2005*.

LINDENBERGER, U., AND P. B. BALTES (1997): "Intellectual Functioning in Old and Very Old Age: Cross-Sectional Results From the Berlin Aging Study," *Psychology and Aging*, 12(3), 410–432.

SINGER, W. (1995): "Development and plasticity of cortical processing architectures," *Science*, 270, 758–764.

SMITH, V., AND J. WALKER (1993): "Monetary Rewards and Decision Costs," *Economic Inquiry*, 31, 245–261.

STANOVICH, K. E., AND R. F. WEST (1998): "Individual Differences in Rational Thought," *Journal of Experimental Psychology: General*, 127(2), 161–188.

TVERSKY, A., AND D. KAHNEMAN (1971): "Belief in the Law of Small Numbers," *Psychological Bulletin*, 2, 105–110.

——— (1982): "Evidential Impact of Base Rates," in *Judgment Under Uncertainty: Heuristics and Biases*, ed. by D. Kahneman, P. Slovic, and A. Tversky, pp. 153–160. Cambridge University Press, Cambridge.

VISCUSI, W. K., AND C. J. O'CONNOR (1984): "Adaptive Responses to Chemical Labeling: Are Workers Bayesian Decision Makers?," *American Economic Review*, 74(5), 942–956.