



The University of Bradford Institutional Repository

<http://bradscholars.brad.ac.uk>

This work is made available online in accordance with publisher policies. Please refer to the repository record for this item and our Policy Document available from the repository home page for further information.

To see the final version of this work please visit the publisher's website. Access to the published online version may require a subscription.

Link to publisher's version: <https://doi.org/10.1016/j.ejor.2017.09.027>

Citation: Goodwin P, Onkal D and Stekler HO (2017) What if you are not Bayesian? The consequences for decisions involving risk. *European Journal of Operational Research*. 266(1): 238-246.

Copyright statement: © 2017 Elsevier. Reproduced in accordance with the publisher's self-archiving policy. This manuscript version is made available under the [CC-BY-NC-ND 4.0 license](https://creativecommons.org/licenses/by-nc-nd/4.0/).



What if you are not Bayesian? The consequences for decisions involving risk

Paul Goodwin*
The Management School
University of Bath
Bath, BA2 7AY, United Kingdom

Email: mnspg@bath.ac.uk
Tel: 01373 461617
Fax: 01373 461617

Dilek Önkal
Faculty of Management & Law
University of Bradford
Emm Lane, Bradford BD9 4JL, United Kingdom
E-mail: D.Onkal@bradford.ac.uk

Herman O. Stekler
Department of Economics
George Washington University
Washington DC 20052, USA
E-mail: hstekler@gwu.edu

*corresponding author

Abstract

Many studies have examined the extent to which individuals' probability judgments depart from Bayes' theorem when revising probability estimates in the light of new information.

Generally, these studies have not considered the implications of such departures for decisions involving risk. We identify when such departures will occur in two common types of decisions.

We then report on two experiments where people were asked to revise their own prior probabilities of a forthcoming economic recession in the light of new information. When the reliability of the new information was independent of the state of nature, people tended to overreact to it if their prior probability was low and underreact if it was high. When it was not independent, they tended to display conservatism. We identify the circumstances where discrepancies in decisions arising from a failure to use Bayes' theorem were most likely to occur in the decision context we examined. We found that these discrepancies were relatively rare and, typically, were not serious.

Keywords: decision processes; Bayes' theorem; judgmental biases; risk

What if you are not Bayesian? The consequences for decisions involving risk

1. INTRODUCTION

A sizeable literature has developed since the 1960s examining the extent to which people depart from Bayes' theorem when they revise subjective probability estimates in the light of new information. The overwhelming conclusion of this research is that people's revised estimates generally do not conform to Bayes' theorem, though the nature of the discrepancies has tended to vary with the type of task being performed. Much of this research has tended to employ tasks relating to book bags and poker chips or quiz-type questions and often the prior probabilities have been supplied by the researcher, rather than being based on the individual's own judgment. While this research has produced many insightful results relating to human cognition and biases, the purpose of making probability judgments is to inform decision making and this literature has generally not considered whether the consequences of departing from Bayes' theorem are likely to be serious when decisions involving risk are being made.

In this paper we address two questions. First, how frequently will decisions based on Bayes' posterior probabilities differ from those based on 'judgmentally-revised' probabilities (we will refer to these as 'discrepant decisions')? Second, when discrepant decisions are made, how serious will the consequences be? We consider these questions in relation to two types of decisions which are commonly encountered: (i) those where the decision involves identifying the most probable of two events, and (ii) those where a decision maker has the option of either receiving an intermediate payoff for certain or a risky choice which will deliver either a better or worse payoff. In both cases, we examine the research questions when the prior probabilities reflect the decision maker's own initial beliefs about the probabilities associated with the decision.

Following a review of the relevant literature, we identify when decisions will be sensitive to departures from Bayes' theorem. We then explore how often discrepant decisions are likely to occur using data from two experiments where decision makers were asked to make probability assessments relating to possible future economic recessions.

2. LITERATURE REVIEW

Assume that a decision maker has a prior probability $p(h_1)$ that an event h_1 will occur so that the implicit prior probability of the event not occurring $p(h_0)$ is $1 - p(h_1)$. If the decision maker revises his or her probabilities according to Bayes' theorem when new information, I , about the state of the world is obtained the revised or posterior probability of the event occurring will be:

$$p(h_1|I) = \frac{p(I|h_1) \cdot p(h_1)}{p(I|h_1) \cdot p(h_1) + p(I|h_0) \cdot p(h_0)} \quad (1)$$

Researchers, particularly economists, have often assumed in their models of human prediction and decision making that people revise their prior beliefs according to (1). For example, Schnader and Stekler (1998) assumed that economic forecasters revise their estimates of the probabilities of recessions according to Bayes' theorem when they receive new information from economic indicators. However, a number of alternative accounts of how people actually revise prior probabilities in the light of new information can be found in the literature. Studies published in the 1960s found that people are conservative in that they do not change their prior probabilities sufficiently when they receive new information when compared with the 'normative' posterior probabilities indicated by Bayes' theorem (Edwards, 1968; Phillips and Edwards, 1966). Later work in areas such as law (Thompson and Schumann, 1987) and accounting (Eger and Dickhaut, 1982) has reported similar findings.

In contrast to the finding that people ‘overweight’ prior probabilities, some studies have suggested that people discount base rates and do not conform to the prior probabilities that they imply. The representativeness heuristic (Kahneman and Tversky, 1973) has been used to explain this. The heuristic appears to be used when the new information is regarded as being more representative of one state of nature rather than others. If this is the case, then that state of nature is given a high posterior probability without reference to the prior probability. The ‘inverse’ fallacy where people confuse $p(I|h_1)$ with $p(h_1|I)$ provides an alternative reason why people might neglect prior probabilities though there has been some debate in the literature as to whether this is distinct from base-rate discounting (Villejoubert and Mandel, 2002).

It is well established that individuals use different heuristics in different decision situations so variations in the nature of the tasks employed in the above studies will probably account for these different findings (Grether, 1992). In this paper our interest is in decisions where people represent their initial beliefs in the form of *their own* prior probabilities. The Bayesian approach to decision making and inference is inextricably linked to subjective ‘degrees’ of belief and how these are revised when new information is available. Our focus is therefore on how a given subjective belief is revised, not on how an externally provided or ‘objective’ probability is revised. Also, we are not concerned with how the prior subjective belief is formed in the first place, though of course, such beliefs will vary between individuals. One of the key influences on the weight that is attached to the prior probability in a given task is likely to be the salience of this probability when compared to other information presented in the task. A prior probability will be expected to have greater relative salience when it is the decision maker’s own estimate. People will be predisposed to pay more attention to their own prior (Evans et al., 2002), especially when it has resulted from some deliberation and effort. Moreover, the ‘advice literature’ suggests that people may have an egocentric bias and attach undue confidence to their own judgments (Bonaccio and

Dalal, 2006). To the extent that new information (e.g. in the form of a forecast) can be regarded as a form of advice, it may be prone to be neglected, particularly if it is perceived to result from statistical analysis rather than a human ‘expert’ (Önkal et al., 2009).

Much research has involved experiments where people were not asked to estimate or even directly consider a prior probability. For example, in the ‘book-bag and poker chip’ studies the experimenters assumed that a given prior probability applied but did not explicitly provide this (Mahmoud and Grether, 1995; Charness et al., 2007; Holt and Smith, 2009). In these circumstances it is perhaps not surprising that the prior probability will be neglected. This means that the discrepancies observed in these studies between the participants’ posteriors and those obtained through Bayes’ theorem may be different from those that we report later.

Another influential factor in the probability revision task is the form in which probabilistic information is provided to forecasters. Some research suggests that information in a frequency form is more likely to be used and used more effectively than information presented as probabilities (e.g. “10 out of every 800 women have breast cancer at age 40” rather than “the probability of breast cancer for a women at age 40 is 0.0125”) (Cosmides and Tooby, 1996; Gigerenzer and Hoffrage, 1995; Koehler, 1996). However there is also some controversy in the literature about what constitutes useful frequency information and why information in this format is useful (see Barbey and Sloman (2007) and associated commentaries). For example, Evans et al. (2000) found that frequency formats were not generally associated with better performance than probability formats when probability information was presented in a way that also allowed accurate perception of the underlying set structure. In particular information on the relative frequency or probability of the conjunction of $I|h_1$ and h_1 (e.g., the relative frequency of occasions when a recession is forecast *and* a recession subsequently occurs) can make the Bayesian computations relatively simple (Barton et al. 2007). Of course, in many important real-world

situations forecasters may not have the opportunity to gather their own frequency information, particularly in relation to the reliability of the new information. Alternatively, their ability to recall this information may be limited so that they have to rely on information presented in summary statistics by a third party which may not be in the convenient conjunctive form.

3. THEORETICAL CONSIDERATIONS

Even though people may not adhere to Bayes' theorem when revising their probabilities the results may not be substantially different from the Bayes' posteriors (Villejoubert and Mandel, 2002; Girotto and Gonzalez, 2001). Recently, Goodwin (2015) suggested two simple heuristics that people could employ to revise prior probabilities and, in most cases, obtain the same decision as that implied by Bayesian revision. However, it is unclear whether people naturally employ these heuristics and, if they do not, whether any deviations from Bayesian revision are likely to be large enough to be consequential when their revised probabilities are used to make decisions. Based on Goodwin (2015) we examine two types of decision: i) decisions that involve the identification of the most probable event (for brevity, we will refer to this as a Type A decision) and ii) decisions involving a choice between obtaining an intermediate outcome for certain or choosing an alternative which will lead to either the best and worst possible outcomes with given probabilities (Type B decision). We use two measures to assess whether a discrepancy between an individual's revised probability and the Bayes' posterior probability is consequential: i) whether the two probabilities imply that different alternatives should be chosen, and ii) the utility loss resulting from the discrepancy. The utility loss is the difference between the expected utility of the best option based on the Bayes' posterior and the 'true' expected utility of the option that would be selected using the individual's revised probability (Barron and Barrett 1996). The 'true' expected utility of this option is calculated using the Bayes' posterior probabilities.

3.1. Type A Decision

In many situations, decision makers face a simple choice which involves identifying which is the most probable of two possible outcomes. This problem often occurs in forecasting when the forecast is to be presented without probabilistic information. For example, a weather forecaster may choose to forecast rain rather than fine weather because s/he judges rain to be more likely (in meteorology these are often referred to as category forecasts). Similarly, in sporting events involving two parties, forecasts of the winner will simply be based on which of the two possible contenders is thought to have the highest probability of victory. This decision can be represented by the simple decision tree in Figure 1 where F1 and F2 are, respectively, forecasts that events E1 and E2 will occur, and w is the probability of event E1. The decision maker's utilities are displayed at the ends of the branches. These are 0 and 1 for the worst and best outcomes respectively. Note that this decision assumes that the utility of choosing the correct option is the same, irrespective of which option is chosen and the same applies to choosing the incorrect option. For example it assumes that an economic forecaster has the same levels of satisfaction with a correct forecast of a recession and a correct forecast of no recession. Let $w = p_B$ if the Bayes posterior probability is used to make the decision after the receipt of new information and let $w = p_E$ if the decision maker's estimate of the posterior probability is used instead. As discussed by Goodwin (2015), the chosen alternatives will differ only if $p_E > 0.5$ when $p_B < 0.5$ or when $p_E < 0.5$

when $p_B > 0.5$. When the decisions do differ, the expected utility loss will be $|2p_B - 1|$.

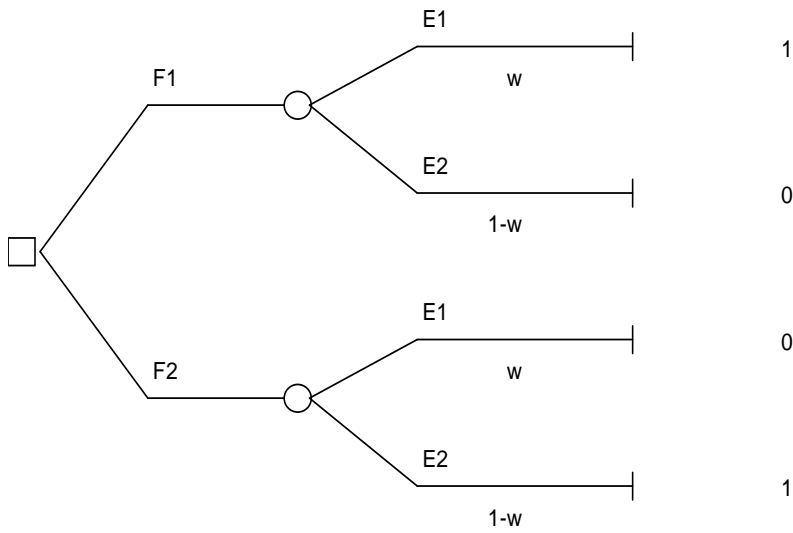


Figure 1
Decision tree for identifying the most probable event (Type A)

3.2. Type B Decision

Figure 2 displays a decision tree for another type of decision problem that people often face either as a complete decision or as part of a more complex decision structure. Here a choice has to be made between an option, A_2 , with a certain intermediate outcome and a risky alternative, A_1 , that will result in either worse or better outcomes, O_1 or O_2 . The utilities of the alternative-outcome combinations are displayed at the ends of the branches and range from 0, for the worst outcome, to 1, for the best, with $0 \leq U \leq 1$. The prior probability of the worst outcome O_1 is x .

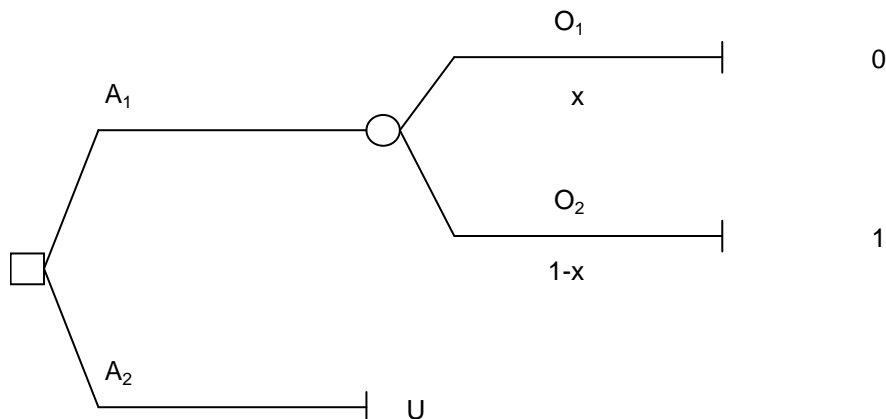


Figure 2
Decision tree for choosing between options with certain and risky outcomes (Type B)

Let $x = p_B$ if the Bayesian posterior probability is used to make the decision after new information has been received and let $x = p_E$ if the decision maker's estimate of the revised probability is used. If the decision maker's preferences are consistent with the axioms of utility theory (e.g., see Goodwin and Wright, 2014), s/he will be indifferent between the alternatives, A_1 and A_2 , when $U = 1 - x$. Suppose that, for a given decision maker, $U = U^*$ and they estimate a posterior probability p_E . Discrepant decisions will be made if:

$$p_E > 1 - U^* \quad \text{when } p_B < 1 - U^*$$

or

$$p_E < 1 - U^* \quad \text{when } p_B > 1 - U^*$$

The expected utility loss of a discrepant decision will be $|U^* + p_B - 1|$.

4. EXPERIMENT 1: SYMMETRIC INDICATORS

We will refer to new information where $p(I|h_1) = 1 - p(I|h_0)$ as a symmetric indicator. This implies that the reliability of the indicator is the same, irrespective of the state of nature. For example, this would apply if a weather forecast has the same probability of being correct whether it is due to rain or not rain. The first experiment was carried out to investigate for the two types of decision referred to above and when an indicator is symmetric: i) the frequency with which a failure to adhere Bayes' theorem will lead to discrepant decisions, and ii) the typical consequences of any discrepancy. The experiment involved participants revising *their own* prior probabilities in the light of new information as part of a decision on whether to invest in an overseas country. We next give details of this experiment.

4.1. Participants

Fifty-six people participated in the experiment. They consisted of 36 MBA students taking a statistics class at Bilkent University (Turkey) and 20 managers taking part in a forecasting workshop at the university. Because of incomplete sets of responses, one participant from each group was deleted. Thus a total of 54 people completed the study. The results from the two groups were combined because they both had experience of working in a business environment (means of 2.4 and 5.9 years respectively) and their overall responses were similar.¹

4.2. Procedure

The participants were told that they had to make nine decisions. In each case, they had to decide whether to invest money in a construction company which was planning to carry out a major project in an overseas country. They were told that the returns on their investment depended on whether the economy in the relevant country continued to grow or whether it would go into a recession. To help them with their decision, they received a brief report on the current state of the country's economy and its immediate prospects, which they were told was compiled by a team of international economists. A total of 27 economic reports were constructed. These were intended to indicate either a high, moderate or low probability of recession. The composite leading indicator indicated that a recession would occur within the next six months (I). Its reliability was either high ($p(I| \text{recession}) = 0.8$) or moderate ($p(I| \text{recession}) = 0.65$), or it provided no useful diagnostic information at all ($p(I| \text{recession}) = 0.5$). If a recession occurred and the decision makers invested, they were told that they would lose \$100,000. The payoff of investing if there was no recession was either \$50,000, \$100,000 or \$200,000. The nine decisions involved reports, suggesting high, medium and low probabilities of

¹ The mean prior probabilities of recession for the two groups were found as 0.422 (MBA) and 0.417 (Managers). The respective standard deviations were 0.081 and 0.083. A t-test showed that the means were not significantly different ($p=0.83$). The participants' mean posterior probabilities of recession (i.e. their mean revised estimates in the light of the economic reports) were 0.530 (MBA) and 0.519 (Managers) with standard deviations of 0.095 and 0.091, respectively. A t-test also showed that the difference between these means was not significant ($p=0.67$).

recession, combined with leading indicators with the three levels of reliability shown above. In addition, the three payoffs each appeared in three decisions, which were randomly combined with the other two conditions. The order of the nine investment decisions was individually randomized for each participant. Eliciting three levels of prior probabilities allowed us to detect any variation in departures from Bayes theorem that might be related to the size of the priors. The use of three levels of payoffs created variety and made the task more realistic. In practice, it would be unusual for a series of nine decisions relating to different countries to have identical payoffs. A lack of variety might also have led to an automated response from participants.

The experiment involved the following stages (a sample response form is given in Appendix A):

- 1) The participants received a brief report on the current and likely future state of the economy for a hypothetical country (the language used in these reports was designed to be similar to those used in practice). They were told that they had been invited to invest a large sum of money in a construction company which is planning to carry out a major project in that country and that the returns on their investment would depend on whether the country's economy continued to grow or whether it goes into recession.
- 2) Based on the provided report they were asked for their prior probability that a recession would occur in this country in the next six months.
- 3) The participants were then provided with the prediction of a composite leading economic indicator along with information on the reliability of this indicator.
- 4) They were then asked for their revised (posterior) probability that there would be a recession in this country within the next six months.

- 5) They were asked to choose between investing and not investing in a situation where their return on investment depended on whether or not a recession would take place.
- 6) If they decided not to invest, the participants were asked to explain why they had made this particular decision and what return they would have required in order to have undertaken this investment.

To provide an incentive for accurate judgment, the participants were told that there were three prizes for those who made the ‘best’ decisions. After the experiment, the total expected payoffs of the decisions made by each participant was calculated (using prior probabilities of 0.7, 0.5 and 0.1 for the high, medium and low probabilities of recession and revising these according to Bayes’ theorem based on the stated likelihoods). Course credits were awarded to the three participants who achieved the highest expected payoffs. Of course, the participants did not know the ‘true’ prior probabilities, but this would be the case in a practical decision making context and the intention was to simulate this as far as possible. The results of stages 5 and 6 will not be reported here. Their purpose was twofold. First, they allowed the probabilities to be elicited as part of a decision process rather than as isolated values which had no subsequent application. Secondly, it enabled us to identify the participants who should be awarded course credits.

5. RESULTS FOR EXPERIMENT 1

5. 1. Did the Decision Makers Conform to Bayesian Revision?

As indicated above, when the economic reports were written, they were designed to encourage the elicitation of three levels of prior probabilities: low, medium and high. Table 1 shows that the manipulation was successful. A general linear model was used to apply a repeated measures ANOVA to the prior probabilities. The differences between the means of the priors that were elicited after participants had read each type of economic report were highly significant.

Intended prior probability	Mean elicited prior probability
Low	0.217
Medium	0.495
High	0.767

(Main effect significant at $p < 0.0005$; $F_{2,106} = 345.0$. All pairwise differences significant at $p < 0.0005$ after Bonferroni correction)

Table 1
Relationship between intended and elicited prior probabilities

We used Bayes' theorem to calculate the posterior probabilities for each participant given their stated prior probabilities and the reliability of the prediction from the economic indicators. These probabilities were then compared with the actual revisions made by the participants themselves. Table 2 shows the mean differences between the two [*i.e.*, Mean(Participants' posteriors – Bayes' posteriors)] for the nine combinations of prior probabilities and levels of reliability. A negative value indicates conservatism while a positive value indicates over-adjustment in response to the economic index.

Intended prior	Mean difference with Bayes' posterior *		
	Reliability of indicator		
	<u>0.5</u>	<u>0.65</u>	<u>0.8</u>
Low	0.14	0.07	0.06
Medium	0.03	-0.02	-0.10
High	-0.06	-0.07	-0.11

* A negative value indicates conservatism

Table 2

Mean differences between Bayes' posterior probabilities of recession and posterior probabilities estimated by participants

It can be seen that when participants' priors were low they tended to over-revise their probability estimates upwards. This was especially true in the case where the index had a reliability of 0.5 and hence provided no useful diagnostic information and the probability should have remained unchanged. High prior probabilities tended to be under revised. The economic indices therefore seemed to make the participants more reluctant to opt for extreme probabilities.

We developed a policy capturing model to try to understand the process used by participants to revise their priors. The model is intended to represent the process of an 'average' participant; clearly the revisions of individual participants will not conform exactly to this model. Given that that the two salient probabilities involved in the revision process will be the participant's prior and the likelihood (i.e. the reliability of the indicator) a plausible proposition is that people's revisions can be modeled simply as a weighted combination of these two values. Similar processes have been observed in time series forecasting (Bolger and Harvey, 1993; Goodwin, 2005; Lawrence and O'Connor, 1992). This suggests the following model.

$$p(\text{recession}|\text{I}) = w_0 + w_1 p(\text{recession}) + w_2 p(\text{I}|\text{recession}) \quad (2)$$

where the w_i are the parameters of the model (we would expect w_0 to be zero).

The model was fitted to the data using the method of generalized estimating equations which is suitable for repeated measures data (Liang and Zeger, 1986). The result is shown below with 95% Wald confidence limits for the parameters displayed below the model in square brackets.

$$p(\text{recession}|\text{I}) = -0.017 + 0.66 p(\text{recession}) + 0.41 p(\text{I}|\text{recession}) \quad (3)$$

[0.57 to 0.75] [0.31 to 0.51]

All p-values for the parameters were less than 0.001 except for the intercept for which $p=0.576$. The point estimates of these parameters, as shown in (3), suggest that, on average, participants' revised probabilities can be approximated by taking a weighted average of the prior and the likelihood with weights of roughly 0.6 and 0.4, respectively. This model gave a better fit to the data than the more sophisticated belief adjustment model proposed by Hogarth and Einhorn (Hogarth and Einhorn 1992) when fit was measured using the Akaike Information Criterion based on quasi-likelihood (QIC) (Pan 2001). The QIC of the simple model was 9.745 while it was 26.881 for the belief-adjustment model; smaller values indicate a superior fit. The simple and belief-adjustment models were also fitted to each individual participant's estimates using ordinary least squares so that parameters specific to that individual could be estimated. Again, the simpler model fits the data more closely. Its mean adjusted R^2 value was 82% compared to 80% for the belief-adjustment model (it should be noted the high adjusted R^2 values will to some extent reflect the relatively small number of observations to which each individual's model was fitted)². Overall it therefore appears that the process typically used by the participants to revise their prior probabilities was a simple one and involved them placing their updated probability estimates between their prior and the supplied likelihood value. This process is very similar to the "Take-the-Average" heuristic suggested by Goodwin (2015). Figure 3 shows, for the case where $p(\text{I|recession}) = 0.8$, how different prior probabilities of a recession would be revised according to the simple model and according to Bayes' theorem.

² When analysing the data for all the participants we used the method of generalized estimating equations (GEE) because of the repeated measures for each participant. This estimation method does not yield an R-squared value, but uses the QIC value instead. The models for each individual participant were fitted using ordinary least squares which does allow R-squared to be measured.

5.2. Consequences of Departures from Bayesian Revision

We next investigate the consequences of departing from Bayesian revision for the two types of decision that were introduced earlier. The decision on whether to invest is a Type B decision, as it involves a choice between a risk-free option and a risky alternative. However, because the probabilities were assessed separately, before the decision was made, we were also able to assess their potential impact on Type A decisions, which would involve determining whether a recession or no recession was most the most probable outcome.

5.2.1. Type A decisions

Recall that for Type A decisions which simply involve identifying the most probable event (see Figure 1), a discrepant decision will only occur if the Bayesian posterior probability is greater than 0.5 when the decision maker's revised estimate is below 0.5 and vice versa. Consider Figure 3, which refers to cases where $p(I|\text{recession}) = 0.8$. Table 2 shows that the largest discrepancies from the Bayesian posteriors are observed for this likelihood. The horizontal line represents values where the revised probability is 0.5. It can be seen that participants' revised probabilities as represented by the simple model (3) are only on opposite sides of this line for prior probabilities between 0.20 and 0.29. This suggests that if people revise their probabilities in accordance with the simple model, then they will make the wrong choice between the least and most probable outcomes relatively rarely. Consistent with this, an analysis of revisions for all participants across all likelihoods revealed that they would have led to the correct choice on 88.8% of occasions (this excludes decisions where either the p_B or p_E values were 0.5 so that the two events E1 and E2 would have been judged to be equally probable). Only 2.5% of decisions were discrepant when the likelihood was 0.5. For likelihoods of 0.65 and 0.8 the percentage of discrepant decisions was

13.6% and 16.5% respectively. Thus, ironically, more reliable information was associated with a greater tendency for a discrepant decision to occur.

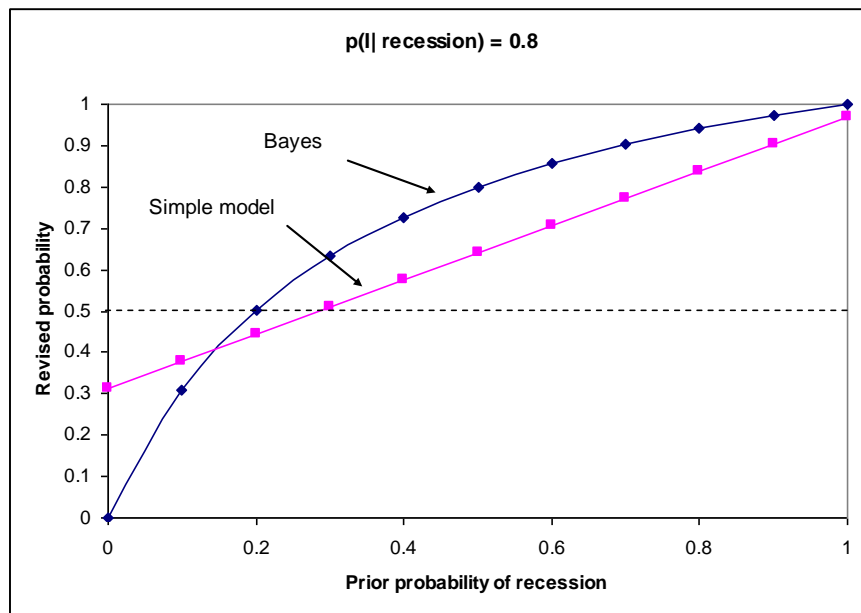


Figure 3
Probability revisions based on Bayes' theorem and simple model when likelihood = 0.8

In the recession forecasting context that we examined, the consequences of a failure to use Bayes' theorem were not serious where Type A decisions are concerned. Recall that, for a Type A decision, the expected utility loss of a discrepant decision is $|2p_B - 1|$. Discrepant decisions had a mean utility loss of 0.40. However, given, that only 45 of the 486 decisions made in the experiment were discrepant, in the vast majority of cases the utility loss was zero. Therefore averaged over all these decisions the mean expected utility loss was only 0.04 (i.e., the mean loss was only 4% of the distance between the worst and best utilities).

5.2.2. Type B decisions

For Type B decisions (Figure 2) the frequency with which a discrepant decision will be made clearly depends on the decision maker's utility (U). We therefore simulated the decisions that would be made by each of our participants if they used either the correct Bayes' posteriors or their own revised probabilities and if their decisions were consistent with each of the following values of U: 0.1, 0.3, 0.5, 0.7 and 0.9. If the payoff of A₂ is halfway between the worst and best payoffs, these utilities will represent attitudes to risk ranging from extreme risk-seeking behavior (0.1) to extreme risk-aversion (0.9) while 0.5 would represent risk-neutrality. Alternatively, they would represent the level of attractiveness of the payoffs of A₂ to the decision maker.

Table 3a shows the percentage of discrepant decisions for high, medium and low values of p_B and the different utilities for A₂. These tended to occur where p_B was low and the utility (U) was high, or vice versa, and Table 3b also shows that discrepancies tended to be more probable when the likelihood was 0.8 showing, as with Type A decisions, that more reliable information is associated with a greater probability of a discrepant decision. The overall values show what percentage of the 486 simulated decisions for each level of utility (i.e. 54 participants x 9 decisions) were discrepant. Depending on the utility, roughly between 5% and 14% of decisions were discrepant with most discrepancies occurring when the utilities did not have extreme values.

Bayes posterior (p_B)	Utility of certain payoff (U)				
	0.1	0.3	0.5	0.7	0.9
below 0.33	0.0	3.2	10.8	29.9	16.6
0.33 to under 0.67	0.0	5.7	15.0	5.7	0.0
0.67 and over	21.2	28.0	3.7	0.5	0.0
Overall	8.2	13.6	9.3	11.5	5.3

Table 3a
Percentage of discrepant decisions for high, medium and low values of p_B

Utility of certain payoff (U)					
Likelihood	0.1	0.3	0.5	0.7	0.9
0.50	3.1	6.2	1.9	11.7	5.6
0.65	3.7	16.7	11.7	9.9	6.2
0.80	17.9	17.9	14.2	13.0	4.3

Table 3b
Percentage of discrepant decisions for different likelihoods

However, the occurrence of a discrepancy does not necessarily imply serious consequences. The mean expected utility losses displayed in Tables 4a and 4b suggest that, on average, the decision makers were unlikely to lose much satisfaction by employing their judgment to revise their prior probabilities rather than Bayes' theorem. For example, the worst mean expected utility loss occurs when the decision makers were assumed to have a utility for the payoff of A_2 of 0.1 and the Bayesian posterior probability (p_B) was over $2/3$. However, even this utility loss is only 12% of the difference between the utilities of the worst and best possible outcomes and most of the other mean expected utility losses were less than 3.5% of this difference.

Utility of certain payoff (U)					
Bayes posterior (p_B)	0.1	0.3	0.5	0.7	0.9
below 0.33	0.000	0.016	0.035	0.044	0.009
0.33 to under 0.67	0.000	0.005	0.013	0.009	0.000
0.67 and over	0.120	0.024	0.010	0.003	0.000
Overall	0.047	0.016	0.019	0.018	0.003

Table 4a
Mean utility losses for high, medium and low values of p_B

Utility of certain payoff (U)					
Likelihood	0.1	0.3	0.5	0.7	0.9
0.50	0.002	0.008	0.005	0.021	0.004
0.65	0.008	0.011	0.020	0.015	0.002
0.80	0.077	0.028	0.031	0.017	0.003

Table 4b
Mean utility losses for different likelihoods

6. EXPERIMENT 2: ASYMMETRIC INDICATORS

We refer to an indicator as asymmetric where $p(I|h_1) \neq 1 - p(I|h_0)$. This may be the case, for example, where an economic indicator is less reliable when a recession is on the way than when no recession will occur. Experiment 2 investigated how people revise their prior probabilities when they are given an asymmetric indicator.

6.1 Participants and Procedure

Fifty-nine final year business students taking a forecasting course at Bilkent University participated in the experiment. The procedure was similar to Experiment 1, though steps 5 and 6 which required participants to make a decision on whether to invest, were omitted. Twenty-seven economic reports were constructed using the same countries and reports as in Experiment 1. Each participant estimated a prior and posterior probability of recession for 9 countries – three with reports suggesting a high prior probability of recession, three suggesting a medium probability and three suggesting a low probability. For each country, an economic indicator was then provided with a level of reliability that was either $p(I|\text{recession}) = 0.8, 0.65$ or 0.5 , as in Experiment 1. However, this time $p(I|\text{no recession})$ was either $0.1, 0.3$ or 0.5 so the indicator was only symmetric for the $(0.5, 0.5)$ combination. For each country, the value of $(I|\text{no recession})$ was randomly assigned to the participants.

7. RESULTS FOR EXPERIMENT 2

7.1 Did the Decision Makers Conform to Bayesian Revision?

As in Experiment 1, the manipulation intended to encourage the elicitation of three levels of prior probabilities was successful, though, where the intended probability was high, the mean elicited prior probability was less than in Experiment 1 (see Table 5).

Intended prior probability	Mean elicited prior probability
Low	0.196
Medium	0.334
High	0.573

(Main effect significant at $p < 0.0005$ $F_{2,116} = 111.9$. All pairwise differences significant at $p < 0.0005$ after Bonferroni correction).

Table 5
Relationship between intended and elicited prior probabilities

As before, we used Bayes' theorem to calculate the posterior probabilities for each participant, given their stated prior probabilities, and the reliability of the prediction from the economic indicators. These probabilities were then compared with the actual revisions made by the participants themselves. Table 6 shows mean the differences between the two [*i.e.* Mean(Participants' posteriors – Bayes' posteriors)] for the combinations of prior probabilities and levels of reliability. As before, a negative value indicates conservatism while a positive value indicates over-adjustment in response to the economic indicator.

Intended Prior	L=0.5			L=0.65			L=0.8		
	p(I no recession)			p(I no recession)			p(I no recession)		
	0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
Low	-0.14	-0.01	0.16	-0.22	-0.04	0.06	-0.28	-0.17	-0.07
Med	-0.30	-0.08	0.05	-0.31	-0.18	0.01	-0.29	-0.09	0.01
High	-0.31	-0.16	-0.02	-0.32	-0.11	-0.05	-0.20	0.01	0.02

Table 6
Mean differences between Bayes' posterior probabilities of recession and posterior probabilities estimated by participants

In contrast to Experiment 1, it can be seen that the participants tended to be conservative. Rather than overreacting to the economic indicator they tend to underweight it, particularly when $p(I| \text{no recession})$ was 0.1. The smaller this value is then the greater the probability of a recession will be as there is a smaller chance of a recession being indicated when none will occur (i.e., the likelihood ratio in favor of a recession is greater). Thus it appears that the participants were unable or unwilling to use the information on $(I| \text{no recession})$ when revising their estimate. As in Experiment 1, we used the method of generalized estimating equations to fit a linear model to the data, this time including $p(I| \text{no recession})$ as an additional independent variable. The result is shown below with p-values for the coefficients displayed below the model in parentheses and 95% Wald confidence limits for the parameters in square brackets.

$$p(\text{recession}|I) = 0.06 + 0.63 p(\text{recession}) + 0.16 p(I| \text{recession}) + 0.07 p(I| \text{no recession}) \quad (4)$$

(0.18)	(0.00)	(0.01)	(0.13)
[0.53 to 0.72]	[0.05 to 0.27]	[-0.04 to 0.17]	

The non-significant coefficient for $p(I| \text{no recession})$ provides further evidence that the participants neglected this information. Moreover, compared to the policy capturing model obtained for the first experiment (3), they attached a much lower weight to $p(I| \text{recession})$ (0.16 versus 0.41). Thus, although the weight attached to the prior probabilities is very similar for the two models (0.63 versus 0.66), it appears that the more complex information relating to the reliability of the economic indicator meant that it had less influence on the revised probabilities. It is possible that the participants found it difficult to comprehend why the probabilities, $p(I| \text{recession})$ and $p(I| \text{no recession})$ did not add up to 100% so they were less trustful of the economic indicator and

discounted it more. This is despite the fact that in most cases it provided a stronger indication that a recession was imminent than in the first experiment.

7.2. Consequences of Departures from Bayesian Revision

7.2.1. Type A decisions

The participants' revisions would have led to the correct decision on 82% of occasions (again this excludes decisions where either the p_B or p_E values were 0.5), only slightly less than the 88.8% rate for experiment 1. However, when $p(I|\text{no recession})$ was equal to 0.1 only 58.5% of decisions were correct, reflecting participants' inability to respond to the high likelihood ratio. In contrast, when this probability was equal to 0.3 or 0.5, 91.3% of decisions were correct. As was the case with the symmetric indicator in Experiment 1, the consequences of a failure to use Bayes' theorem do not appear to be serious where Type A decisions are concerned, at least in the context we explored. Averaged over all 486 decisions the mean expected utility loss would only have been 0.07 (i.e. the mean loss was only 7% of the distance between the best and worst utilities). Recall that it was 0.04 when the indicator was symmetric.

7.2.2. Type B decisions

As before, we simulated the decisions that would be made by our participants if they used either the correct Bayes' posteriors or their own revised probabilities and if their decisions were consistent with each of the following values of U : 0.1, 0.3, 0.5, 0.7 and 0.9. Table 7a shows the percentage of discrepant decisions for high medium and low values of p_B and the different values of U . As before, these tended to occur where p_B was low and the utility (U) was high or vice versa. Table 7b shows the percentage of discrepant decisions for high medium and low values of the

likelihood ratio and different values of U. As expected, given the conservatism shown by the participants and the neglect of the values of $p(I| \text{no recession})$, discrepant decisions would be more common the greater the value of the likelihood ratio. However, this was not the case when U is low.

Bayes posterior (p_B)	Utility of certain payoff (U)				
	0.1	0.3	0.5	0.7	0.9
below 0.33	1.9	1.9	4.5	23.6	26.8
0.33 to under 0.67	5.2	5.2	30.5	26.4	2.3
0.67 and over	24.0	60.0	29	8.5	0.5
Overall	10	24.9	22.2	18.8	8.9

Table 7a
Percentage of discrepant decisions for high medium and low values of p_B

Likelihood ratio	Utility of certain payoff (U)				
	0.1	0.3	0.5	0.7	0.9
<2	7.5	17.3	11.7	14.5	13.1
2 to 3	13	20.4	16.7	15.7	8.3
≥ 5	14.4	34.9	35.9	24.9	4.8

Table 7b
Percentage of discrepant decisions for different likelihood ratios

Tables 8a and 8b show that, as with a symmetric indicator, on average decision makers are unlikely to lose much satisfaction by employing their judgment to revise their prior probabilities rather than Bayes' theorem. The maximum value recorded in the two tables is 0.089. However, even this utility loss is only 8.9% of the difference between the utilities of the worst and best possible outcomes and most of the other mean expected utility losses were less than 5% of this difference.

Utility of certain payoff (U)					
Bayes posterior (p_B)	0.1	0.3	0.5	0.7	0.9
below 0.33	0.000	0.008	0.014	0.026	0.015
0.33 to under 0.67	0.008	0.010	0.027	0.046	0.006
0.67 and over	0.013	0.080	0.089	0.042	0.003
Overall	0.008	0.036	0.046	0.038	0.008

Table 8a
Mean utility losses for high medium and low values of p_B

Utility of certain payoff (U)					
Likelihood ratio	0.1	0.3	0.5	0.7	0.9
<2	0.010	0.022	0.022	0.022	0.010
2 to 3	0.006	0.030	0.031	0.018	0.004
≥ 5	0.007	0.053	0.079	0.065	0.008

Table 8b
Mean utility losses for different likelihood ratios

8. DISCUSSION AND CONCLUSIONS

Our study differs in three important aspects from those reported in earlier research into the extent to which people's revised probabilities conform to Bayes theorem. First our tasks involved the revision of participants' own subjective probabilities in the light of new information. Secondly the revision was set within the context of a decision. Thirdly, the consequences for decisions of any departures from the probabilities indicated by Bayes theorem were assessed. In general, we found that people tended to overreact to a symmetric indicator when their prior probability of a recession was low and underreact when it was high. This led to a reluctance to estimate extreme posterior probabilities and may have arisen because they used a heuristic that involved averaging their prior probability with $p(I|recession)$. In contrast, they tended underreact to an asymmetric indicator and hence were conservative in their revisions. The discounting of the new information in

this case may have resulted from the more complex information about the indicator's reliability which many participants may have had difficulty in absorbing.

Despite these biases, we found that discrepant decisions resulting from departures from Bayes' theorem were relatively rare whether the indicator was symmetric or asymmetric. Moreover, the consequences of these errors, when assessed in terms of the mean loss of utility, were typically far from serious. If these results could be generalized beyond the immediate context of our experiments (we discuss reservations as to whether this would be the case below) then they would have a number of implications. For example, they would suggest that models which attempt to predict the choices of decision makers on the assumption that they do revise their probabilities according to Bayes' theorem would still be accurate even though this assumption is unlikely to be valid (e.g., Schnader and Stekler, 1998). Second, they would suggest that in decision support systems an emphasis should be placed on the elicitation of well calibrated prior probabilities in the first place, rather than Bayesian revision of these probabilities. Imposing Bayes-revised probabilities on decision makers may be unnecessary and problematical for those who seek ownership of their decision processes and regard the Bayes formula as a black box.

Of course, our findings may be dependent on the specific characteristics and context of the task that we employed. It involved people revising *their own* prior beliefs about probabilities in order to make decisions that will be similar to those that they will face in their professional lives. Nevertheless, our experiments only allowed us to explore the effect of indicators that were (with the exception of where the likelihood ratio was 1) designed to encourage prior probabilities to be revised upwards. In addition, setting the revision task within the context of a decision carries some dangers. In particular, there is evidence that forecasts can be biased by the relative attractiveness of potential outcomes of the decision (Weber, 1994; Fildes et al., 2009). For example, people

sometimes confuse forecasts with decisions that will subsequently be based on them and bias their forecasts in the direction of the most desirable outcomes (Krizan and Windschitl, 2007). The specific context of this experiment -forecasting recessions - may also have had a particular influence on the results we obtained (Holt and Smith 2009). It is likely that participants will have had prior beliefs about the reliability of economic indicators and it is possible that skepticism about the accuracy of macro-economic forecasts may have distorted their use of the supplied likelihoods. However, again there is no evidence that the economic indicator forecasts were ignored. Indeed, as we've seen, when people had low prior probabilities and the indicator was symmetric they tended to overreact to the forecasts based on the indicators.

The presentation of the likelihoods as percentages is another factor that potentially limits the extent to which inferences can be drawn from our results. These percentages provided no direct information on the relative frequency of the conjunction of events (a recession occurring *and* a forecast of recession) which can help to simplify the Bayesian computations. Also the percentages were presented to participants while in many situations they will be learned over time as part of an experiential process. However, as we argued earlier, these limitations may be typical of many real-world situations and given that the impact of departures from Bayes' theorem in our experiment were rare and inconsequential, a removal of this limitation would be expected to lead to only marginal reductions in the frequency of discrepant decisions.

Finally, our task imposed a decomposition structure on the decision makers. The elicitation of prior probabilities revised probabilities and the decisions to invest was achieved through separate questions which had to be answered sequentially. Our conclusions might not therefore be applicable in situations where risky decisions are made intuitively and where all information

relating to the problem is considered holistically and simultaneously rather than separately and serially.

The structures of our decision were relatively simple but they are likely to be typical of many decisions that people encounter. Promising extensions would be to further investigate how people perform on tasks where there are more than two possible outcomes, where the outcomes are measured on a continuous scale (e.g. future world temperatures) or where multiple rather than single sources of new information are provided (Cameron, 2005). In addition, further work could usefully examine the effect of indicators that encourage prior probabilities to be revised downwards (e.g. indicators that predict that a recession will not occur) and assess the extent to which our findings can be generalized to other contexts beyond economic forecasting.

REFERENCES

- Barbey, A.K. & Sloman, S.A. (2007). Base-rate respect: from ecological rationality to dual processes. *Behavioral and Brain Sciences*, 30, 241–297.
- Barron, F.H. & Barrett, B.E. (1996). Decision quality using ranked attributes weights. *Management Science*, 42, 1515-1523.
- Barton, A., Mousavi, S., & Stevens, J.R. (2007). A statistical taxonomy and another “chance” for natural frequencies. *Behavioral and Brain Sciences*, 30, 255-256.
- Bolger, F. & Harvey, N. (1993). Context sensitive heuristics in statistical reasoning. *Quarterly Journal of Experimental Psychology*, 46A (4), 779-811
- Bonaccio, S. & Dalal, R.S. (2006). Advice taking and decision making: an integrative literature review and implications for organizational sciences. *Organizational Behavior and Human Decision Processes*, 101, 127-151.

- Cameron, T.A. (2005). Updating subjective risks in the presence of conflicting information: an application to climate change. *Journal of Risk and Uncertainty*, 30, 63–97.
- Charness, G., Karni, E., & Levin, D. (2007). Individual and group decision making under risk: an experimental study of Bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and Uncertainty*, 35, 129-148.
- Cosmides, L. & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58, 1–73.
- Edwards, W. (1968) Conservatism in human information processing. In: B. Kleinmuntz B (Ed.), *Formal Representation of Human Judgment* (pp. 17-52). New York: Wiley.
- Eger, C. & Dickhaut, J. (1982). An examination of conservative information processing bias in an accounting framework. *Journal of Accounting Research*, 20, 711-723.
- Evans, J. St. B.T., Handley, S.J., Over, D.E., Perham, N., & Thompson, V.A. (2000). Frequency versus probability formats in statistical word problems. *Cognition*, 77, 197- 213.
- Evans, J. St. B.T., Handley, S.J., Over, D.E., & Perham, N. (2002). Background beliefs in Bayesian inference. *Memory and Cognition*, 30, 179-190.
- Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25, 3-23.
- Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, 102, 684–704.
- Giroto, V. & Gonzalez, M. (2001). Solving probabilistic and statistical problems: a matter of information structure and question form. *Cognition*, 78, 247-276.
- Goodwin, P. & Wright, G. (2014). *Decision Analysis for Management Judgment*. Fifth Edition. Chichester: Wiley.

- Goodwin, P. (2005). Providing support for decisions based on time series information under conditions of asymmetric loss. *European Journal of Operational Research*, 163(2), 388-402.
- Goodwin, P. (2015). When simple alternatives to Bayes formula work well: Reducing the cognitive load when updating probability forecasts. *Journal of Business Research*. 68, 1686-1691.
- Grether, D.M. (1992). Testing Bayes' rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior and Organization*, 17, 31-57.
- Hogarth, R.M. & Einhorn, H.J. (1992). Order effects in belief updating: the belief-adjustment model. *Cognitive Psychology*, 24, 1-55.
- Holt, C.A. & Smith, A.M. (2009). An update on Bayesian updating. *Journal of Economic Behavior and Organization*, 69, 125–134.
- Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.
- Koehler, J.J. (1996). The base rate fallacy reconsidered: descriptive normative and methodological challenges. *Behavioral and Brain Sciences*, 19: 1-.
- Krizan, Z. & Windschitl, P.D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin*, 133, 95–121.
- Lawrence, M. & O'Connor, M. (1992). Exploring Judgemental Forecasting. *International Journal of Forecasting*, 8, 15-26.
- Liang, K-Y. & Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- Mahmoud, A.E. & Grether, D.M. (1995). Are people Bayesian? Uncovering behavioral strategies. *Journal of the American Statistical Association*, 90, 1137-1145.

- Önkal, D., Goodwin, P., Thomson, M., Gönül, M.S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, 22, 390-409.
- Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics*, 57, 120-125.
- Phillips, L.D. & Edwards, W. (1966). Conservatism in simple probability inference tasks. *Journal of Experimental Psychology*, 72, 346-357.
- Schnader, M.H. & Stekler, H.O. (1998). Sources of turning point forecast errors. *Applied Economics Letters*, 5, 519-521.
- Thompson, W.C. & Schumann, E.L. (1987). Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior*, 11, 167-187.
- Villejoubert, G. & Mandel, D.R. (2002). The inverse fallacy: An account of deviations from Bayes's Theorem and the additivity principle. *Memory and Cognition*, 30, 171-178.
- Weber, E.U. (1994). From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin*, 115, 228-242.

APPENDIX A: SAMPLE FORM FOR A COUNTRY

A decision on whether to invest: Tegoland

You have been invited to invest a large sum of money in a construction company which is planning to carry out a major project in the Asian country of Tegoland. The returns on your investment will depend on whether the Tego economy continues to grow or whether it goes into a recession.

Given below is a brief report on the state of the Tego economy by a team of international economists:

“Retail sales in Tegoland declined in most of the last 9 months as a bleaker economic outlook damped consumer demand according to the Bloomberg retail purchasing managers index monthly survey out today.

Record oil prices have pushed up fuel costs crimping household spending. Retail sales will tumble this quarter amid worsening economic conditions.

The global outlook is darkening by the day and Tegoland’s low potential growth rate makes it particularly vulnerable.....Risks to growth are skewed to the downside and even the chance of a recession cannot be dismissed lightly.”

1. Estimate the probability that the Tegoland economy will go into a recession in the next 6 months
--

A composite leading economic indicator of the Tego economy has declined in the last three consecutive months (October to December). In the past:

- in years when a recession occurred the indicator had declined in the previous October November and December on 80% of occasions
- in years when no recession occurred the indicator had previously declined in each of these months on 20% of occasions

2. Use this information to estimate a revised probability that the Tegoland economy will go into a recession in the next 6 months
--

You have to decide whether to invest the money in the construction company.

- If you decide to invest and there is a recession then you will lose \$100000
- If you decide not to invest and there is no recession you will have missed out on an opportunity to make a return of \$100000

3. Do you want to invest in the company? Yes/No

4. If your decision is NOT to invest: a) Please briefly explain why you were not prepared to invest b) What return would you require in order to undertake the investment? \$.....
