Master Thesis Behavioural Economics

Erasmus School of Economics | Erasmus University Rotterdam

Bayesian reasoning or heuristics?

The effect of time pressure

Abstract – Classic economic theories assume that agents make decisions under the beliefs formulated through Bayes' theorem. However, literature in behavioural economics found that agents used various heuristics to form beliefs. What strategy do agents truly use in Bayesian tasks and what triggers decision makers to change their behaviour? This thesis tested the effect of time limitation on the behaviour of decision makers in Bayesian tasks. We found time pressure decreases the use of Bayesian reasoning and triggered decision makers to switch to the quicker but less accurate heuristics as decision rule. However, the switch did not decrease the average accuracy in the Bayesian tasks due to the low number of Bayesian subject.

Author	J.P.J. van Gangelen
Student Number	511715jg
Supervisor	Y. Xu
First reader	Dr. T. Wang
Date	August 18, 2019

Table of contents

1 | Introduction

“Doctors are surprisingly bad at reading lab results. It’s putting us all in a risk”. This is the opening of a news article in the Washington Post on October 5th, 2018 (Morgan, 2018). Doctors especially fail to interpret false-positive predictive values, caused by failure in the use of Bayes’ theorem. Research has shown that the majority of medical students, house staff and physicians overestimate a laboratory test result (Manrai, Bhatia, Strymish, Kohane, & Jain, 2014; Casscells, Schoenberger, & Graboys, 1978). Such failures in Bayesian reasoning suggest potentially tragical consequences. According to Brase and Hill (2015) this is a serious, real-world problem.

Bayes’ Theorem or Bayes’ rule is a mathematical formula (Figure 1.1) for calculating conditional probabilities, named after the English statistician Thomas Bayes (Bayes, 1764). This basic normative model of information processing is used in economic analysis (Holt & Smith, 2009).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad\qquad .075 = \frac{.01 * .8}{.01 * .8 + .99 * .1}$$

Figure 1.1: Bayes' Theorem (left); Bayes’ theorem applied in mammogram-case (right).

A basic Bayesian problem typically contains two levels of information: the base-rate or prior information and the diagnostic or indicant information (Bar-Hillel, 1980). To clarify the theorem, the use will be demonstrated with an example. Suppose, 1% of the female population has breast cancer, this information is called the base-rate. A positive mammogram has a hit rate of 80% and a false alarm rate of 10%. This information is called the diagnostic information. Eddy (1982) asked physicians to estimate the probability that a woman with a positive mammogram actually has breast cancer. In the experiment, 95% of the physicians estimated the probability that she has breast cancer to be between 70% and 80%. Whereas Bayes’ theorem gives a probability of 7.5% (Figure 1.1). This conditional probability is called the posterior probability. A person who acts in line with Bayes’ theorem is called Bayesian. This example shows the importance of correctly combining the base-rate with the diagnostic information. The group of women without breast cancer is much bigger than the group with breast cancer. As a result, the number of false alarm tests is bigger despite the relatively low probability of a false alarm.

Bayes’ theorem is pervasively used in economics to describe how an economically fully rational decision maker processes information and forms beliefs. In practice, people do systematically deviate from traditional economic theories of decision-making such as Bayes’ rule. This behaviour can often be explained by heuristics, which can make cognitive biases and economic non-optimal tendencies predictable (Frederiks, Stenner, & Hobman, 2015). Heuristics are efficient cognitive processes to make decisions more quickly or frugally that, conscious or unconscious, ignore part of the information (Gigerenzer & Gaissmaier, 2011). A more accessible but limited way to describe heuristics is to name them mental shortcuts or rule of thumbs (Shah & Oppenheimer, 2008). Gigerenzer and Todd (1999) described the mind as an adaptive toolbox with various heuristics, tailored for specific social and physical environments. The use of heuristics will often result in systematic errors or biases (Tversky & Kahneman, 1974). Heuristics do not help to optimize the decision, rather try to find an option that exceeds an aspiration level. It is impossible to fully analyse each decision in daily life; people do not have unlimited time, mental effort and cognitive ability. Heuristics help to simplify search and decision problems, which allow people to process information in a less effortful way than one would expect from an optimal decision rule (Shah & Oppenheimer, 2008).

Not every decision is important enough to spend a load of time and energy, people rather accept a loss in accuracy. People make, conscious or unconscious, an accuracy-effort trade-off when facing a decision. People base this trade-off on their decision environment, such as the available time, amount of information and importance of decision (Mousavi & Gigerenzer, 2014). This thesis focuses on the effect of the available time, an important element in decision-making. Time constraints play, to a greater or lesser extent, a role in nearly all decisions in daily life. This time constraint can be the trigger for decision makers to deviate from statistical rules such as Bayes' theorem. The current study tries to clarify the behaviour of decision makers when facing a Bayesian task. Do decision makers act in line with the theorem? Do decision makers use heuristics? And most importantly, what is the effect of available time on their behaviour? Based on these questions the following research question is conducted for this thesis:

How and why does time pressure affect peoples' Bayesian reasoning performance?

This research question will be answered based on experimental research. The extensive literature review will map the potential behaviour of decision makers when solving a Bayesian task. An experiment will provide quantitative data on the Bayesian reasoning performance of decision makers, with and without time pressure. The difference in the performance of participants enables categorization in behavioural patterns. The experiment will generate valuable insight into Bayesian reasoning performance, behaviour patterns and the effect of time pressure.

This paper has academic as well as practical relevance. The research will contribute to a better understanding of individual decision processes, specifically in Bayesian task. The thesis helps to explain how, why and when decision makers' actions differ from Bayes' theorem. The role of heuristics seems to play a central role in this explanation. Understanding these biases can help to design better mechanisms to predict behaviour (Levin, Peck, & Ivanov, 2016). Most importantly, this thesis helps to clarify the effect of time pressure on decision-making processes. This is interesting because the quality of decision-making depends heavily on time (Svenson, Edland, & Slovic, 1990). Time pressure can be a trigger for decision makers to switch from the optimal decision strategy to other decision strategies. The research results will show to what extent the time pressure causes the deviation from optimal behaviour in Bayesian tasks. By better understanding people's economic behaviour, it is possible to inform, help or manipulate people more efficiently. Time pressure in combination with Bayes' theorem is not researched before, research on this combination is unique. Results of this thesis can be applied to areas with social or economic relevance. A better understanding of decision-making under uncertainty and heuristic can be useful in the consumer products market, policy areas, insurances, voting or medical considerations (Grether, 1980).

The remainder of this study is laid out as follows: Chapter 2 will start by reviewing the behavioural patterns in Bayesian tasks, followed by studying the effect of time pressure on Bayesian reasoning performance. Section 3 describes the experimental design and procedure. Chapter 4 shows both the descriptive statistics as well as the results of the analysis. The final chapter, chapter 5 discusses the results of this study and suggests possibilities for future research.

2 | Literature review

This chapter starts by discussing the most presumable behavioural patterns in Bayesian tasks, followed by a review of the literature on time pressure. Hypotheses will be formulated based on the literature.

2.1 Behavioural patterns in Bayesian tasks

When people face a Bayesian task, multiple behavioural strategies can be followed. People do not always act in line with the traditional theory, this can be a conscious or unconscious decision based on the accuracy-effort trade-off. Decision makers can act in line with the normative strategy and try to find the optimal decision or follow strategies which are less accurate and less effortful. Those strategies are used to find a decision that satisfices instead of optimizing (Gigerenzer, 2008). Based on literature six behavioural patterns are distinguished.

2.1.1 Bayesian reasoning

According to traditional economic theory, fully rational individuals would use Bayes' theorem when facing a Bayesian task. With Bayesian reasoning, a decision maker can calculate the posterior probability and make a statistically optimal decision (Achtziger & Alós-Ferrer, 2013). The decision maker will use all crucial information and combines the base-rate and diagnostic information in the correct way. This method to evaluate information seems to be the most logical strategy when facing a Bayesian task. El-Gamal and Grether (1995) confirmed that Bayes' rule is the most used strategy when facing Bayesian tasks. Contrary, Kahneman and Tversky (1972) claimed that people are not acting in line with Bayes' rule in the evaluation of information. Kahneman and Tversky stated that people are: "not Bayesian at all." (p. 450).

In this paper, it is hypothesized that intrinsic or extrinsic motivated decision makers will use Bayesian reasoning to make an optimal decision. Under the condition, the subject has the ability to make the correct calculation. Peoples' ability to solve Bayesian tasks correctly is positively correlated with their intelligence (Furnham & Chamorro-Premuzic, 2004). Secondly, experience is positively correlated with learning (Yelle, 1979). Subjects who had seen the same Bayesian problem before are more likely to give the correct response (Grether, 1980). Therefore, familiarity with Bayesian tasks is assumed to be positively correlated with the ability to solve those tasks correctly.

The mental effort used to solve the task will be rewarded with a better outcome. According to classic theory, economic fully rational people will try to optimize their reward. This Behavioural pattern will result in high accuracy in Bayesian reasoning. Therefore, it is hypothesized that people do use Bayes' theorem when facing a Bayesian task with no time pressure.

H1: People act in line with Bayes' theorem when answering a Bayesian task with no time pressure.

2.1.2 Base-rate fallacy

An alternative behavioural pattern followed by decision makers is to focus on particular information in a Bayesian task. Decision makers ignore or under weigh the base-rate and compose their answer based on the diagnostic information, rather than integrate all crucial information. This is called base-rate fallacy or base-rate neglect (Bar-Hillel, 1983; Compte & Postlewaite, 2018). In general, people do observe the base-rate but chose to ignore the information. Decision makers do this because the base-rate seems to them irrelevant for the judgment that they are making. Base-rates will be taken into account when the two levels of information are perceived as being equally relevant (Bar-Hillel, 1980). The experience of the relation between the base-rate and the diagnostic information determines the use of the base-rates in Bayesian tasks (Christensen-Szalanski & Beach, 1982).

Base-rate fallacy can partly be explained by representativeness heuristic. Representativeness is the heuristic in which people determine their decision by the degree to which an option is similar to its parent population (Kahneman & Tversky, 1972). People's judgement, of the probability that X is a Y, is mediated by the degree to which X is similar to Y (Kahneman & Frederick, 2002; Kahneman & Tversky, 1973). A surprising number of mathematically sophisticated individuals shows the same probabilistic misinterpretation as mathematically naïve individuals. Even when subjects personally witnessed their illogic behaviour in solving problems, they often keep using representativeness heuristic (Cox & Mouw, 1992). Taffler (2010) proved investors use this heuristic, motivated subjects stimulated by monetary incentives. This heuristic can cause large biases, especially in the situation of conjunction error. People judge a joint event to be more likely than is one of its elements considered alone (Wells, 1985). In a Bayesian task, the diagnostic information can be used to decide which option is more representative (El-Gamal & Grether, 1995; Grether, 1992). The diagnostic information in combination with representativeness heuristic can explain biased behaviour of decision makers. El-Gamal and Grether (1995) found this base-rate fallacy as the second most important behavioural pattern when facing a Bayesian task. Contrary, Koehler (1996) argues that base-rates are almost always used and often influence judgments in reasonable ways. The degree of use depends on task presentation and structure.

The base-rate fallacy will in most cases decrease the accuracy of the posterior probability estimation compared to Bayesian reasoning, since the base-rate information is crucial for optimal reasoning performance. On the other hand, this method will save time and mental effort. People can choose to prefer less effort over accuracy.

2.1.3 Base-rate anchor

Decision makers can use anchor-and-adjustment heuristic when facing a Bayesian task. When using the anchor-and-adjustment heuristic people make an estimation by starting from an initial value, the anchor, which is adjusted to generate the final answer (Tversky & Kahneman, 1974). The anchor-and-adjustment heuristic tends to be insufficient because people stop adjusting as soon as the nearest point in a range of plausible values is reached (Epley & Gilovich, 2006). People adjust differently upwards than downward, the effect of a higher anchor is significantly larger than lower anchors (Jacowitz & Kahneman, 1995). Ariely, Loewenstein and Perlec (2003) showed the mechanism of this heuristic by asking people whether they would buy a good for a dollar figure equal to the last two digits of their social security number. After their yes/no response, participants were asked to give their highest willingness to pay for the product. Participants with above-median social security numbers were willing to pay values from 57 to 107 percent greater than subjects with below-median numbers. The participants anchor on their social security number.

A second explanation for anchoring-and-adjustment heuristic is the activation of anchor-consistent knowledge. When a person tests whether an anchor is plausibly it activates the anchor-consistent knowledge which affects the final answer (Wegener, Petty, Blankenship, & Detweiler-Bedell, 2010). Anchor-consistent knowledge comes easier to mind. This is closely related to availability heuristic; decision makers evaluate the probability of events by the ease with which relevant cases come to mind (Tversky & Kahneman, 1973). People’s individual differences affect the strength of anchoring-and-adjustment heuristic. The anchoring effect is related to intelligence, the anchoring effect decreases with higher cognitive ability (Bergman, Ellingsen, Johannesson, & Svensson, 2010). Secondly, people with experience in the task are less affected by the provided anchors (Wilson, Houston, Etling, & Brekke, 1996). Surprisingly, motivated participants often fail to adjustment better from provided anchors compared to unmotivated participants (Tversky & Kahneman, 1974; Chapman & Johnson, 2002). Epley and Gilovich (2005) found a decrease in the anchor effect for a self-generated anchor when participants had higher accuracy motivation.

In Bayesian tasks, decision makers can use the base-rate as an anchor. In most cases, this task-related information is the first crucial information decision makers face. The information can be used as an externally provided anchor. People use the diagnostic information to adjust from the anchor or will ignore the diagnostic information completely. Numerical anchors influence just about any type of judgment. Research illustrates the robustness of the heuristic both in and outside the laboratory (Furnham & Boo, 2011). Base-rate anchoring is the third most prominent rule El-Gamal and Grether (1995) found in their research on behaviour in Bayesian tasks. This strategy is closely related to the conservatism effect, the underestimation of the impact of new evidence (Tversky & Kahneman, 1974). Subjects give too much weight to the base-rate and need more evidence to change their believes than prescribed by Bayes' theorem (Philips & Edwards, 1966).

The incorrect way of evaluating the information will lead to biases and decrease the accuracy of the posterior probability estimation. However, the anchoring-and-adjustment heuristic will cost less mental effort and the less complicated process will save time. Decision makers have limited time and cognitive capabilities. Therefore, people can prefer the use of heuristics over Bayesian reasoning, despite the decrease in Bayesian reasoning performance.

2.1.4 Prior posterior probability anchor

When subjects face multiple Bayesian tasks a fourth behavioural pattern can be distinguished. This behaviour is once more based on anchoring-and-adjustment heuristic. Decision makers can use the posterior probability of a previous task as an anchor. In this strategy, people will make a calculation for the first Bayesian task, the starting point, and adjust this value in the following tasks. The prior posterior probability will be used as a self-generated anchor. A self-generated anchor is processed differently than anchors provided by an experimenter. Self-generated anchors can be quickly adjusted since the accuracy of the anchor is known; the anchor does not need to be evaluated as an externally provided anchor. People give answers closer to self-generated anchors (Epley & Gilovich, 2001). Moreover, the anchor is plausible in most cases. Extreme anchors have a weaker effect compared to plausible anchors (Wegener, Petty, Blankenship, & Detweiler-Bedell, 2010). Every numeric value can be used as an anchor. However, a self-generated and plausible values will have a stronger effect.

Gehlbach and Barge (2012) research anchoring-and-adjustment in questionnaire responses. Anchoring occurs in questionnaires between adjacent items with related content. Decision makers use their response to an initial item and adjust from that anchor. The information on the new task is taken into account, but the adjustment is typically not sufficient (Joyce & Biddle, 1981). In a questionnaire, adjacent items invite people to use anchoring-and-adjustment heuristic more compared to non-adjacent items (Gehlbach & Barge, 2012). When people see a pattern in the adjacent task, the use of the previous answer is more likely.

This behavioural pattern is more efficient but will lead to less accurate Bayesian reasoning performance, because of the bias caused by anchoring. It is practically impossible to optimize each decision in daily life; this will cost too much time and cognitive resources. Therefore, decision makers can decide to process not all the task information sufficiently.

2.1.5 Incorrect statistical rule

The use of an incorrect statistical rule is the fifth behavioural strategy. Decision makers tend to calculate the joint probability ($P(A \cap B)$) instead of the conditional probability ($P(A \mid B)$) when facing Bayesian tasks. Pollatsek, Well, Konold, Hardiman, and Cobb (1987) researched the understanding of conditional probabilities. In their research, they claim it is likely that some subjects' mistakes are related to confusion between the joint and conditional probabilities of the events. To clarify the difference between those two methods an example (Salop, 1987) will be used. Suppose Rotterdam has two types of taxis 85% green taxis and 15% red taxis, an eyewitness can identify the correct colour of the taxi 80% of the time. When an eyewitness reported a taxi red, what is the probability the taxi is truly red? What people tend to do is calculate the probability a taxi is correctly identified and red, which is $P(A \cap B)=$ 80%*15%. However, people should calculate the probability a taxi is correctly identified given the taxi is red. This should be calculated with Bayes' rule: $P(A \mid B)=$ (80%*15%)/((80%*15%)+(20%*85%)). The probability that Events A and B both occur is called the probability of the intersection of A and B. Interpretation of the conditionality as an intersection event is identified as one of the common mistakes in students' thinking processes when calculating a conditional probability (Huerta & Lonjedo, 2007). In their research, a substantial part of the students answer questions about a conditional probability by means of a joint probability.

People can use this behavioural strategy out of convenience. This method is easier and will cost less time and mental effort. As a matter of course, using an incorrect statistical method will decrease the accuracy in solving Bayesian tasks. Decision makers' trade speed and effort for accuracy. On the other hand, people can use the joint probability method out of ignorance. People may have difficulty with the structure of conditional probability statements, performance in those tasks depends on the details of wording (Pollatsek et al., 1987). The decision maker could think he/she solves the Bayesian task correctly with this joint probability method or does not master the use of Bayes' theorem.

2.1.6 Random

People do always make trade-offs between costs and benefits. When the costs of Bayesian reasoning or heuristic are too high compared to the incentives, decision makers can decide to give random answers. Randomly answering will cost minimal effort and is used when there is low intrinsic and extrinsic motivation. Another reason for random answers is the lack of cognitive ability (Camerer & Hogarth, 1999). In this case, the lack of cognitive ability required for the Bayesian task. People can misunderstand the task or simply do not know how to combine the base-rate and diagnostic information. This behavioural pattern will lead to unpredictable decisions. As a matter of course, random answers will lower the accuracy of the decision.

The last bias to consider is peoples' tendency to overestimate low probabilities and underestimate high probabilities when facing a Bayesian task. This bias is found by Holt and Smith (2009) in their research on Bayesian updating.

2.2 Time pressure

Decision-making is affected by time pressure, the quality of decision making depends heavily on time (Svenson, Edland, & Slovic, 1990). Under time pressure people have the feeling the given time is insufficient to perform their decision-making process or the given time is factually insufficient to execute the process. Time pressure often increases the level of stress, pressure and arousal. This will affect people decision-making processes (Maule & Hockey, 1993). Findings of Maul, Hockey and Bdzola (2000) showed that time-pressured participants were more anxious and more energetic.

Research on the effect of time pressure on forced decision-making defined three general ways in which people can respond to time constraints (Dhar & Nowlis, 1999; Weenig & Maarleveld, 2002; Ben Zur & Breznitz, 1981). First, decision makers increase the rate at which they examine information. People will still evaluate all the information but increase the speed of reading and evaluating the information. Examining information faster can result in less accurate examination. When facing time pressure, the decision maker may not be able to continue adequate control over the processing of all pieces of information. With a possible decrease in performance at a certain level of information load (Hahn, Lawson, & Lee, 1992). Secondly, decision makers tend to filter information more rigorous such that they can focus on the more important information. People place more weight on important information compared to decisions without time pressure. This is an effective but less precise way to make a decision, not all information is taken into account. Ignoring information that should be taken into account will lower accuracy. Decision makers will base their decision on incomplete information. Evaluating less evidence before making a decision will result in less accurate but faster decisions (Donkin, Little, & Houpt, 2014). Thirdly, decision makers may change their decision strategy. A common response to limited time is the use of less effortful decision strategies. People simplify their strategy, use a strategy which is feasible with incomplete information. For example, shift from using compensatory to non-compensatory decision rules (Svenson, Edland, & Slovic, 1990). These faster strategies will often deviate from the optimal decision. Decision makers can be forced by the time limit to alter their strategy or chose to do so due to the feeling of pressure. All three ways to respond to time constraints can and often will lower the accuracy of the decision-making process. This is an example of a dilemma generally known as speed-accuracy trade-off. In many situations, people have to negotiate between the demand for response speed and response accuracy (Bogasz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010). This ability to trade accuracy for speed is fundamental in human decision-making (Donkin, Little, & Houpt, 2014).

Time pressure has been shown to reduce the quality of decision-making (Payne, Bettman, & Johnson, 1993; Edland & Svenson, 1993). McDaniel (1990) concluded that auditors' processing accuracy declined as time pressure increased and audit efficiency increased with increasing time pressure. Working capacity is believed to moderate the effect of time pressure. People with high working capacities would perform better than people with low working capacities when processing higher information loads under time pressure (Hahn, Lawson, & Lee, 1992). It is interesting to see the effect of time pressure on peoples' behaviour, decision-making under time pressure is a part of many peoples' daily lives (Ahituv, Igbaria, & Sella, 1998).

2.3 Time pressure and behaviour in Bayesian tasks

In this research, the main question is how time pressure affects people's behaviour in Bayesian tasks. The effect of time pressure on Bayesian reasoning has not been studied before. Nevertheless, by combing literature on time pressure and the determined behavioural patterns in Bayesian tasks, substantiated hypotheses can be composed for this deductive research. It is assumable that behaviour will change in two ways. First, part of the people will keep following the same behavioural pattern. Their strategy will not change but due to the time limit, they will be less precise in their calculations. People will round numbers more easily when using Bayesian reasoning or adjust less accurate from an anchor. This will lower the accuracy of Bayesian reasoning performance. Secondly, people will change their behaviour as a result of the time limitation. People will choose a behavioural pattern that cost less time effort, this will infer a shift to the more time-efficient heuristics or randomization. Time pressure increases the use of heuristic processes (Goodie & Crooks, 2004), especially when people have high motivation to process information (Suri & Monroe, 2003). Decision makers shift from the slow and optimal decision process to a faster and more autonomous process. As explained, this is a trade-off between accuracy and effort. These faster behavioural patterns will cost less time, but this will negatively affect the accuracy of the Bayesian reasoning performance. The use of heuristics will often result in systematic errors, people ignore or do not sufficiently process all crucial information of the tasks. Heuristics cannot be expected to match performance under the more optimal Bayesian reasoning.

Time pressure will affect the accuracy in Bayesian task performance. This time limitation will presumably cause changes in behavioural patterns as well as a change in preciseness within the processes. Both changes will lower the accuracy in Bayesian reasoning. Therefore, it is hypothesized that time pressure will decrease the accuracy in Bayesian tasks.

H2: Bayesian reasoning performance decreases under time pressure.

As explained, time pressure will presumably change the behavioural patterns used by decision makers. People think or factually do not have enough time to act in line with Bayesian reasoning when facing a time limit. The time pressure will exert pressure on decision makers' information evaluation. People want to or must use other strategies to answer the Bayesian task. Therefore, it is hypothesized that time pressure will decrease the use of Bayesian reasoning.

H3: The use of Bayesian reasoning in Bayesian tasks decreases under time pressure.

Decision makers who do not use Bayesian reasoning can use one of the heuristics or chose to give a random answer. A shift from Bayesian reasoning to heuristics is more likely since people who would use Bayesian reasoning with no time pressure are motivated to give an accurate answer. Subjects who are willing to act Bayesian in situations without time pressure are assumable willing to invest the smaller effort which is needed for the heuristics in the situation with time pressure. Therefore, it is hypothesized that time pressure will increase the use of heuristics.

H4: The use of heuristics in Bayesian tasks increases under time pressure.

To conclude, it must be noted that the use of heuristics in specific circumstances do not lower the performance under time pressure. In exceptional situation, heuristics not only reduce effort but even increase accuracy. For example, when optimization is computationally intractable (Gigerenzer, 2008). Bobadilla-Suarez and Love (2018) tested decision heuristics under time pressure and concluded more frugal heuristics are not necessarily faster than less frugal ones. In parallel, less frugal strategies can be fast given the right context. Whether a heuristic is fast and/or frugal depends on the type of heuristic and the situation. The faster heuristic can be more accurate than a slower strategy, therefore time pressure can possibly increase the Bayesian reasoning performance. Additionally, Goodie and Crooks (2004) challenged the assumption that time pressure decreases performance in decision tasks by increasing the use of heuristics. They suggest heuristics can be surprisingly effective when they are used in their optimal environment. The behaviour of participants was tested in a probability-learning procedure. Time pressure did not systematically decrease performance and even improved performance in one case, heuristic processes are not always suboptimal.

3 | Methodology

For this thesis, an experiment is conducted in order to generate quantitative data on this specific research topic. With this methodology, it is possible to control variables, vary factors of interest and generate specified data (Croson & Gächter, 2010). In this section, the research design and selected research methods will be discussed.

3.1 Measure of Bayesian performance

The Bayesian performance is measured in a series of Bayesian tasks. The Bayesian tasks are based on the experiment of Grether (1980) in his research on the effect of representativeness heuristic on violation of Bayes' rule. Participants face three virtual boxes: Box X, box A and box B. Box X contains ten balls marked with an 'A' or a 'B', the number of 'A' and 'B' differ in each task. Box A contains ten balls, n red balls and 10-n blue balls. Box B contains ten balls, n red and 10–n blue. The procedure starts with randomly selecting a ball in box X. The letter on this selected ball determines whether the next ball is drawn from box A (ball with 'A') or B (ball with 'B'). The participant does not know whether box A or B is selected. Thereafter, the computer randomly selects a ball from the selected box. This red or blue ball is showed to the participant. The participant is asked to access the probability of the ball is drawn out of box A. In this Bayesian task, the information on the number of balls with an 'A' or 'B' in box X is the base-rate information. The number of red and blue balls in box A and box B is the diagnostic information. Participants are asked to assess the posterior probability in terms of the chance out of 100, as in the research of Holt and Smith (2009). Asking participants to give the posterior probability in an open question generates a richer set of choices, which makes it easier to distinguish behavioural patterns when analysing the data. In general, participants have trouble to give a probability in exact numbers. However, the straightforward design of these mathematical tasks and the benefits of this scale outweighs this disadvantage. For comparison, in previous research more difficult tasks were used, but participants only had to choose whether box A or B was more likely (El-Gamal & Grether, 1995; Grether, 1980; Achtziger & Alós-Ferrer, 2013). The posterior probability in this experiment can easily be translated into binary answers (box A P>50%, box B P<50%).

The experiment consists of six tasks. This is comparable to the number of tasks in the research of Kahneman and Tversky (1973) who tested the behaviour of their subjects in five tasks. Without any time or cost limitations, it would be preferred to test the behaviour of subjects in multiple more tasks, which would make the statistical test more reliable. In this research the potential reward is too low, subjects are not willing to spend too much time and effort in the experiment.

Each of the six tasks differs in base-rate and diagnostic information, see Table 3.1 and Figure 3.1. When designing the Bayesian tasks the goal was to create tasks which most people are able to solve, at the same time the task should not be too easy. Subjects can solve the task, but is it worth the mental effort? The values are carefully selected to distinguish behavioural patterns and to observe the behaviour in both low and high probability outcomes. The posterior probabilities in these six tasks are used to calculate the Bayesian reasoning accuracy and recognize behavioural patterns. Hypotheses 1 can be tested with this information.

Table 3.1: Base-rate, diagnostic information and posterior probability of the six Bayesian tasks.

	Base-rate	**Diagnostic information**		**Posterior probability**
#	*Box X (a/b)*	*Box A (r/b)*	*Box B (r/b)*	*Red ball, box X*
1	2/8	4/6	9/1	10%
2	4/6	2/8	4/6	25%
3	7/3	3/7	7/3	50%
4	5/5	9/1	6/4	60%
5	4/6	6/4	1/9	80%
6	9/1	9/1	1/9	99%

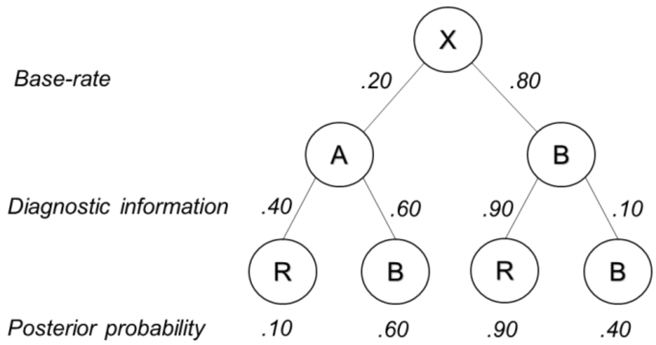


Figure 3.1: Diagram posterior probability Bayesian task 1.

3.2 Time limitation

The participants are divided into two groups, one treatment group and one control group. Both groups will face the exact same tasks, the only difference is a time limit for each Bayesian task in the treatment group. The participants are randomly assigned to one of the treatments, randomization protects against selection bias (Angrist & Pischke, 2009). This control over the experiment makes causal inferences more reliable.

The treatment group has a time limit for each of the six Bayesian tasks. This time limit is based on a pre-test, executed to determine the time subjects need to perform the task when there is no time constraint. This pre-test is identical to the control treatment in the final experiment. The only difference, subjects had the chance to win € 10.- in the pre-test. Following the procedure suggested by Benson and Beach (1996), the time constraint for the experiment is set at one standard deviation below the pre-test subjects' mean time. In this case, this will be based on the mean of subject who act in line with Bayes' theorem (see Table 3.2). In the pre-test, data of twelve Bayesian-subjects was collected. This is comparable to the number of subjects used in the pre-test of Suri & Monroe (2003), who tested ten subjects in there pre-test to determine time constraints. The pre-test showed a clear difference in the time needed for each of the tasks. Subjects need more time for the first task. It is assumable this is because subjects need to understand the task and choose a strategy to solve the task. Secondly, there is a time difference between the tasks which presumably caused by the variance in difficulty of the calculations.

Table 3.2: Mean and standard deviation of time in pre-test & time limitation (in seconds).

#	Mean	Standard deviation	Time limit
1	126.75	45.43	81
2	68.33	19.43	49
3	34.58	9.60	25
4	52.42	28.12	24
5	34.00	11.55	22
6	58.67	23.86	35

Participants should be able to finish the task with Bayesian reasoning within this time. However, the time will create pressure and there is no time for much hesitation. If times are normally distributed, this forced about 84 percent of the Bayesian participants to execute the task faster than normal. Participants who are able to execute the task faster will feel the pressure of time constraint (Ordonez & Benson, 1997). With this between-subject design with two treatments, the effect of time pressure on Bayesian reasoning performance and behavioural patterns can be observed. The measures of this treatment effect are needed to test hypotheses 2, 3 and 4.

3.3 Measures of cognitive ability, motivation, experience and time

The ability to solve a statistical task, the effect of time pressure and the use of heuristic is correlated with the participant's cognitive ability (Bergman et al., 2010; Furnham & Chamorro-Premuzic, 2004:

Hahn, Lawson, & Lee, 1992). Therefore, the cognitive ability of the participants is tested and used as a control variable. A comprehensive cognitive ability test is not suitable for this experiment due to time and effort limitations of the participants. In this experiment the Cognitive Reflection Test (CRT) is used, this simple measure is introduced by Frederick (2005). The CRT measures participants' ability to reflect on a question and resist the first response that comes to mind. The test is based on the distinction between two types of cognitive processes; the rapid and autonomous process versus the slower and reasoning process (Evans & Stanovich, 2013). The test consists of three items, see Appendix A. Frederick (2005) compared de CRT with other, more extensive, measures and found positive significant correlations. The disadvantage of the CRT is the familiarity of participants with this measure. Stieger and Reips (2016) showed that 44 percent of the participants had experiences with these tasks. Participants with prior exposure to the task score substantially higher CRT scores (Haigh, 2016).

The motivation of the participants and experience in Bayesian reasoning can affect the task performance (Yelle, 1979), the effect of time pressure (Suri & Monroe, 2003) and the use of heuristic (Epley & Gilovich, 2005; Wilson et al., 1996). It is important to control for this factor, in order to make the correct conclusion on the variable of interest. Therefore, the participants are asked to indicate their motivation to answer the Bayesian task correctly. A 7-point scale (Strongly agree-Strongly disagree) is used as in research on motivation by Vallerand et al. (1993). The experience in Bayesian reasoning is measured by asking the participants whether they are familiar with similar tasks (yes/no). Subjects who had seen a similar Bayesian problem before are more likely to give the correct response (Grether, 1980).

Furthermore, the response time is recorded this can help to identify the behavioural patterns and the effectiveness of the treatment. Bayesian reasoning is a controlled and slower proces, it will cost more time compared to the use of heuristics or a random guess. Heuristics are based on an automatic process and will result in a quick response (Achtziger & Alós-Ferrer, 2013).

3.4 Demographic question

The experiment contains four demographic questions. This will give some information on the participants and indication of the generalizability of this study. The participants are requested to give their gender in a multiple-choice question (female, male, other) and their age in an open question. Thirdly, the participants are asked what their highest level of completed education is (8 options) and whether they have learned statistical or mathematical courses after high school (yes/no). The demographic will not be used as a control variable as there is no indication these factors will affect the tested behaviour. Grether (1980) argued in his paper that conventional economic theories do not discriminate among types of individuals.

3.5 Monetary incentive

A task-related payment is offered, this monetary incentive makes this experiment an economic experiment. Incentives give control and people are stimulated to use their costly mental effort (Read, 2005). In this experiment binary lottery incentives are used, subjects can earn lottery tickets in each task

(Harrison, Martínez-Correa, & Swarthout, 2014). The number of lottery tickets is linearly correlated with the correctness of their answers, which induce saliency (Smith, 1982). Participants receive four lottery tickets for the correct posterior probability and one ticket less for each two-percentage points deviation from the optimum. With this method, giving the correct answer is optimal for the participant, but a substantiated guess can also be lucrative. It stimulates participants to give a well-considered answer. After the experiment, one ticket will be randomly selected, this participant earns €30. Monetary incentives seem to increase the accuracy in Bayesian task of experienced subjects (Grether, 1980). Contractionary, research showed that monetary incentives for accuracy did not reduce the effect anchoring heuristic (Tversky & Kahneman, 1974) or increase the accuracy of subjects who were unfamiliar with Bayesian tasks (Grether, 1980). The monetary incentive should stimulate the participants to answer the tasks correctly. At the same time, the incentive should not kill all non-Bayesian behaviour.

3.6 Procedure

The experiment is executed online. An online experiment is the most suitable method for this research, taken time and cost limitation into account. The disadvantage of an online experiment is the lower control over cofounding factors, the environment of the participants is variable. Qualtrics is the tool used to design and perform the experiment. For a copy of the experiment see Appendix B.

Participants are asked to read instructions before starting the experiment. In the instructions the participants are explained their privacy is guaranteed, the experiment is anonymous. This leads to higher validity of data; participants will behave more natural. Participants will not be driven by social values as pride or fear (Pearlin, 1961). Secondly, there will be an indication of the duration of the experiment; approximately 8 minutes. As well as a statement to explain the data is only used for this master thesis for the Erasmus School of Economics and a few words of thanks for participating in the research. The instruction part will end with an explanation of the possibility to earn money in the experiment.

After the instruction page, participants start the experiment. First, the design and procedure of the tasks will be explained. Participants will be asked to read this information carefully before they start with the first task. The six Bayesian tasks will be tested, each task on a separate page. Participants are asked to give their answer in rounded numbers. There will not be a default option since this can nudge the participants’ answers (Dinner, Johnson, Goldstein, & Liu, 2011). It is only possible to continue the experiment after filling in an answer. The treatment group will face a clock on the bottom and top of the screen, which indicates the second left until time ran out. If a participant failed to give an answer within the time limit, a choice screen will be shown and the participant is asked to decide immediately (Maule, Hockey, & Bdzola, 2000). The participant can win a maximum of two lottery tickets for the correct answer if answered after the time run out. The higher the deviation from the correct answer, the lower number of tickets the participant will earn. This method is used to motivate subjects to give the answer within the time limit in the next tasks. At the same time, subjects are motivated to give a well-considered answer when time run out which can give valuable insights.

After finishing the Bayesian tasks, participants will face the measures of cognitive ability, motivation and experience. Lastly, the demographic question will be tested. After completing the experiment, participants have the opportunity to fill in their email address to participate in the lottery. This email address will only be used to elect the winner of the monetary reward.

4 | Results

This chapter comprises the presentation and analysis of the findings resulting from the experiment. Based on the quantitative analysis of data, the hypotheses are tested. In total 194 participants started the experiment, 129 of these participants completed the experiment. Which entails a relatively high dropout rate of 33.5 percent. Nearly all the respondents who did not finish the experiment quitted the experiment in the introduction phase or during one of the first tasks. Both the treatment and control group are equally represented in the sample; 63 participants completed the experiment in the treatment group and 66 participants in the control group. The output of each statistical test used in this chapter can be found in Appendix C.

4.1 Bayesian reasoning performance

The six Bayesian tasks were completed by 129 participants, this resulted in 774 data points to analyse the Bayesian reasoning performance. Figure 4.1 shows the average answer and median of each task, combined with the correct posterior probability. Notable is the deviation from the optimal answer. The average answer is, in nearly all tasks, substantially lower compared to the optimal answer.

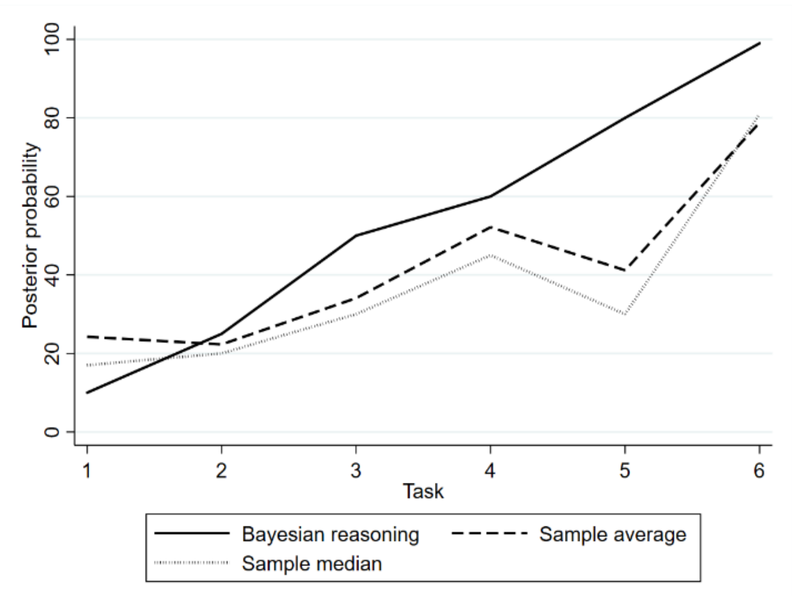


Figure 4.1: Sample average and sample median of posterior probabilities answered by participants compared to Bayesian optimal.

Participants' Bayesian reasoning performance is based on the absolute deviation from the posterior probability. The higher the deviation, the lower the performance. A participant who acts perfectly in line with Bayesian reasoning will have zero deviation, which is the highest possible score. The mean total performance of the sample is -127.55, with a standard deviation of 56.90. The 75th percentile of the total Bayesian reasoning performance is -104.

4.1.1 Treatment effect

Without time limitation people score on average -120.50. This is higher than the average Bayesian reasoning score in the treatment group, in which people score -134.94. The two groups are compared with the Mann-Whitney U test, suitable to compare two non-parametric samples with each other. The test proves there is no significant (10%) difference between the rank of the two treatments. In other words, this test proves that there is no significant difference in Bayesian reasoning performance under time pressure.

The average performance in each task for the treatment group and the control group are graphically presented in Figure 4.2. In four out of the five tasks, people perform on average better in the control group compared to the treatment group. Participants in both groups have clearly lower Bayesian reasoning performance in the fifth task. Only in task 1 the difference in distribution between the control and treatment group is significant (5%).

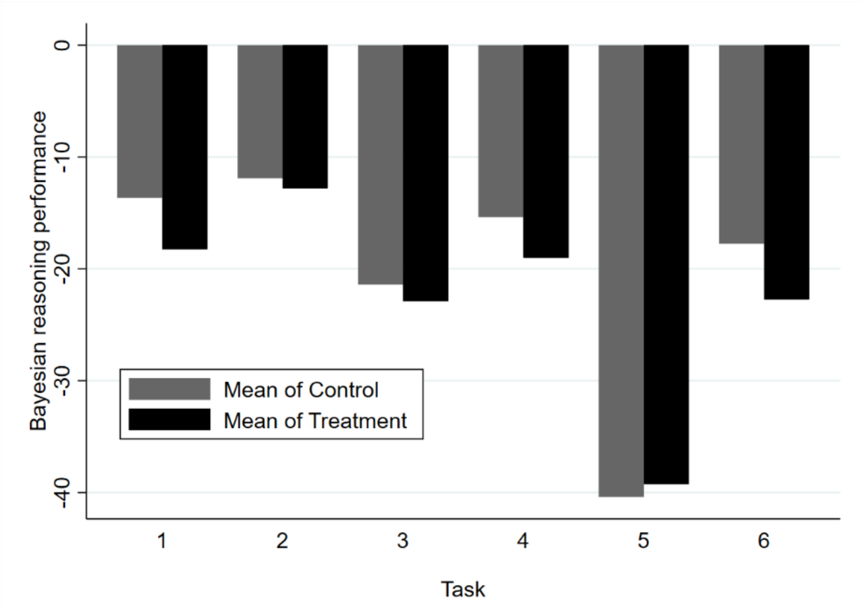


Figure 4.2: Average Bayesian reasoning performance in each task (1-6) for the treatment group and control group.

The mean Bayesian reasoning performance, the average time used for each task and the standard deviation of the sample can be found in Table 4.1. On average the treatment group spent 124.62 seconds and the control group 234.63 seconds to finish the six tasks. In the first tasks, there is a large difference in mean time between the control and treatment group. The difference between treatment and control groups is smaller in the last tasks. Remarkable is the difference between the mean time and the time limits in the treatment group. This difference is large in most tasks, people do not use all the time available.

Table 4.1: Bayesian reasoning performance (score) per task; time in seconds used to finish the task; Standard deviations (SD); time limitation treatment group.

	Control		Treatment		
#	**Score (SD)**	**Time (SD)**	**Score (SD)**	**Time (SD)**	**Time limit**
1	*-13.65 (18.74)*	*99.19 (158.18)*	*-18.25 (20.53)*	*46.50 (21.06)*	*81*
2	*-11.91 (7.41)*	*45.42 (58.23)*	*-12.79 (6.86)*	*23.60 (13.18)*	*49*
3	*-21.41 (12.45)*	*32.39 (39.57)*	*-22.90 (10.40)*	*15.32 (7.85)*	*25*
4	*-15.38 (11.89)*	*21.11 (18.27)*	*-19.02 (13.02)*	*14.12 (7.04)*	*24*
5	*-40.39 (22.46)*	*19.42 (18.27)*	*-39.24 (20.82)*	*12.68 (6.59)*	*22*
6	*-17.76 (17.20)*	*17.10 (15.59)*	*-22.73 (23.65)*	*12.40 (7.41)*	*35*
Total	*-120.50 (58.53)*	*234.63 (219.35)*	*-134.94 (54.62)*	*124.62 (48.57)*	236

In task 4, 5 and 6 the average time used in the control group is below the time limitation in the treatment group. This indicates that the time limitation only worked for the first three tasks. In the first three tasks, the treatment group used a significant (5%) lower amount of time compared to the control group, in the last three tasks this difference is not significant (10%). To get a better understanding of this possible difference in treatment effect, the Bayesian reasoning performance in the first and last three tasks between and within the groups are compared. There is no significant (10%) difference between the performance in the first three tasks between the control and treatment group. Nor a significant (10%) difference between the performance in the last three tasks between control and treatment group. So, the treatment effect is not significantly different in the second part compared to the first part of the tasks. However, there is a difference in performance within the samples. Participants' performance decrease in the second part of the experiment. Both within the control and within the treatment group there is a significant (1%) decrease in performance between the first three and last three tasks.

4.1.2 Linear regression

It is hypothesised Bayesian reasoning performance decreases under time pressure. This hypothesis will be tested with a multiple linear regression. The independent variables are the discrete variables; Cognitive ability, Motivation, and dummy variables; Familiarity and Treatment. The continuous dependent variable is the Bayesian reasoning performance. With this model, it is possible to see the effect of the time limitation on Bayesian reasoning performance controlled for variables which are known to influence the performance.

The cognitive response test determined the cognitive ability of the sample. Frederick (2005) divided the CRT-scores into three categories; "Low" if zero out of the three questions were answered correctly, "Intermediate" for one or two correct answers and "High" for three correct answers. In this sample, most participants are classified in the intermediate-group (49%), second-most in the high-group (37%) and only 14 percent in the Low-group. Secondly, the participant's motivation. In general, the participants were motivated to answer the tasks correctly; 26 percent was strongly motivated, 36 percent motivated and 21 percent slightly motivated. Only 9 percent of the participants were slightly unmotivated, unmotivated or strongly unmotivated. It is plausible unmotivated subjects quit the experiment in an early stage, the motivation of these dropouts is not measured. Thirdly, the participants' experience. The majority (69%) was not familiar with comparable tasks, 31 percent was familiar with the task. Cognitive ability and experience are equally distributed in the treatment and control group. However, there is a significant (5%) difference in the distribution of motivation in the two groups. Participants who have faced the time limitation indicated a significant (5%) lower motivation.

The output of the linear regression is presented in Table 4.2. A higher cognitive ability score, motivation or experience does not increase the Bayesian reasoning performance significantly (10%). The expected effect of these variables is not visible in the data. Being part of the control group compared to the treatment group does not increase the Bayesian reasoning performance significantly (10%). This outcome shows once again that there is no significant decrease in Bayesian reasoning performance under time pressure. The R-squared of the model is 0.0671, which means only 6.71 percent of the variation in Bayesian reasoning is explained by the variables. This low percentage in combination with the insignificance of the variables indicates the weaknesses of this model.

Two additional linear regression were executed, one with the Bayesian reasoning performance in the first three tasks as dependent variable and one with the performance in the last three tasks as dependent variable. This is done because of the significant difference between the first and second part of the experiment, differences in Bayesian reasoning performance and time used to solve the tasks. These differences can affect the treatment effect. However, in both regressions the Treatment variable was not significant (10%) and the R-square did not improve substantially.

Table 4.2: Multiple linear regression outcome for the relationship between the treatment and Bayesian reasoning performance; controlled for cognitive ability, motivation, time, and experience.

	Task 1-6	Task 1-3	Task 4-6	Task 1-6
Cognitive Ability	**3.462 (0.488)**	3.184 (0.155)	0.277 (0.943)	3.618 (0.463)
Motivation	**-5.617 (0.148)**	0.743 (0.667)	-6.359* (0.034)	-4.406 (0.258)
Experience	**11.269 (0.333)**	8.987** (0.085)	2.282 (0.799)	12.187 (0.290)
Treatment	**11.158 (0.264)**	5.990 (0.179)	5.168 (0.501)	
Time				0.061* (0.042)
R^2	**0.0671**	0.0688	0.0556	0.0885

*Note. Number of observations = 774. P-value between brackets; * significant 5% level. ** significant 10% level.*

The fifth column in Table 4.2 shows the output of a regression where the variable Treatment is replaced by the variable Time. There is a significant (5%) relation between the time used to answer the tasks and the Bayesian reasoning performance. When the participant's time used to answer the six tasks increases with one second the Bayesian reasoning performance increase with 0.061 points, ceteris paribus.

4.2 Behavioural patterns

Based on the literature, six behavioural patterns were identified: Bayesian reasoning, base-rate fallacy, base-rate anchor, prior posterior probability anchor, incorrect statistical reasoning, and random. The expected answers for each behavioural pattern can be found in Appendix D. These theoretical answers are compared with the answers given by the respondents.

4.2.1 Identification rules

The first two patterns, Bayesian reasoning and base-rate anchoring, are identified with a model introduced by Grether (1980) and discussed in detail by Benjamin (2018). This model (Figure 4.3) shows to what extent an answer is based on the base-rate and diagnostic information. The p refers to the true probabilities in the task and the π is the decision maker‘s answer. The c correspondents to the use of diagnostic information and the d to the use of the base-rate. If a decision maker acts perfectly in line with Bayes’ theorem, the diagnostic information and base-rate will be used correctly (c=d=1). A c or d smaller than 1 corresponds to underweighting, larger than 1 to overweighting the information. When d=1 and c=0, the answer is purely based on the base-rate. A logistic regression (Figure 4.2) is used to calculate the use of diagnostic information (β_1) and base-rate (β_2).

1)
$$\frac{\pi\ (A|S)}{\pi\ (B|S)} = \left[\frac{p(S|A)}{p(S|B)}\right]^c \left[\frac{p(A)}{p(B)}\right]^d$$

2)
$$ln\left(\frac{\pi(A|S)}{\pi(B|S)}\right) = \beta_0 + \beta_1\, ln\left(\frac{p(S|A)}{p(S|B)}\right) + \beta_2 \ln\left(\frac{p(A)}{p(B)}\right)$$

Figure 4.3 (1) Model to identify the use of diagnostic information (c) and base-rate (d).
Figure 4.4 (2) Logistic regression to identify the use of diagnostic information (β_1) and base-rate (β_2).

In the experiment conducted for this thesis only π (A│S) is measured, the π (B│S) is assumed to be 1- π (A│S). This is assumable in case of Bayesian reasoning (p(B│S) = 1-p(A│S)) and base-rate anchoring (p(B) = 1-p(A)). It is not possible to identify base-rate fallacy with this method because in this experiment p(S│A) ≠ 1-(S│B). For each participant, an individual regression is executed. The average R-squared in the sample is 0.725, in most regressions a large part of the variance can be explained by the diagnostic information and base-rate. When people act perfectly in line with Bayesian reasoning or base-rate anchoring the R-squared is 1. A participant who used incorrect statistical reasoning has a fit of 0.873. In this research, a participant can be assigned to Bayesian reasoning or base-rate anchoring category when the R-squared is bigger than or equal to 0.95. This relatively high R-squared limits the chance to falsely classify participants to a behavioural pattern. At the same time, this limit allows a participant to deviate from the pattern to a certain extent. When the R-squared is bigger or equal to 0.95, the coefficients are used to identify the correct pattern. The average β_1 is 0.449 and the average β_2 is 0.515. A participant is marked as Bayesian when both coefficients are between 0.7-1.3 and a participant's behaviour is identified as base-rate anchoring when only β_2 is between 0.7-1.3. This range has been chosen to allow the participant to deviate, to a certain extent, from the theoretical pattern. The range must allow the participant to answer one of the six tasks not in line with the pattern and/or make small miscalculations. At the same time, this range filters regressions with a high R-squared but extraordinary coefficient.

The next three behavioural patterns are both types of base-rate fallacy and the use of incorrect statistical reasoning. These patterns are identified based on the extent of homogeneity between the collected data and the theoretical answers of these behavioural patterns. In this research, it is assumed participants follow a behavioural pattern when five of the six tasks were answered within an error-margin. This allows the participant to answer one task random or totally wrong. For each answer of the participant, the absolute deviation from the theoretical answers is calculated. The five tasks which are the closest to one of the patterns are used to classify the participant behaviour. These five answers can deviate jointly 15 percentage points from a behavioural pattern to be assigned to the category. The diagnostic information and base-rate values in the six tasks are selected such that it is impossible for a participant to be classified in more than one pattern. Multiple error-margins were tested, the selected range gives the participant some margin for error and is at the same time unlikely to identify a behavioural pattern falsely. The margin is in line with margins in the first identification rule. When this error-margin was used to identify Bayesian reasoning or base-rate anchoring, nearly the same number of participants was classified to those categories.

For the last behavioural pattern, prior posterior probability anchoring, a different method is used. In theory, a subject who use this heuristic makes a calculation in the first task and adjust this value in the following tasks. In this experiment, the participants should increase their answers in each adjacent task. The behaviour of participants is classified as 'prior posterior probability anchor' when the following three criteria were met. 1) An increase in answered values in five or six adjacent tasks. 2) No other behavioural pattern identified. 3) Maximum of 60 percentage point deviation from the Bayesian optimum.

Subjects who did not meet the criteria of one of the behavioural patterns is classified as random. To check whether no behavioural patterns were overlooked, all subjects of this 'Random' category were compared to each other. None of the participants in this category has answered more than four tasks with the same values compared to another subject in this category. Only two patterns with four overlapping values were found, two participants met those two patterns. This indicates a good qualification of behavioural patterns.

4.2.2 Identified behavioural patterns

In Table 4.3 the identified behavioural patterns can be found, for each treatment group and for the sample in total. A behavioural pattern is identified for 59.69 percent of the subjects.

Table 4.3: Number of participants for each behavioural pattern; specified for the treatment group, control group, and the total sample.

	Treatment	Control	Total
Bayesian reasoning	*2*	*9*	*11*
Base-rate anchor	*6*	*4*	*10*
Base-rate fallacy 1	*7*	*3*	*10*
Base-rate fallacy 2	*2*	*0*	*2*
Prior posterior anchor	*2*	*1*	*3*
Incorrect statistical rule	*15*	*26*	*41*
Total	*34*	*43*	*77*
Random	*29*	*23*	*52*
Total	*63*	*66*	*129*

There are several numbers which catch the eye when looking at Table 4.3. First of all, the remarkable low number of participants in the Bayesian reasoning category, only 11 in total. The first hypothesis stated that people act in line with Bayes' theorem when answering a Bayesian task with no time pressure. The most conservative way to test this hypothesis is to determine whether a majority of the participants in the control group acts in line with Bayes' rule. Only nine of the 66 participants in the control group act in line with the theorem. Obviously, this is below the 50 percent boundary. The binominal test and proportion test confirm that it is extremely likely (99% confidence level) that the number of Bayesian people in the population is below 50 percent. People do not act in line with Bayes' theorem when answering a Bayesian task with no time pressure.

The percentage of Bayesian subject is low in the situation with no time pressure, but even lower in the situation with time pressure. The third hypothesis stated that the use of Bayesian reasoning decreases under time pressure. Fisher's exact test is used to analyse whether the difference in proportion between two binominal groups is significant. The test proves that the proportion of Bayesian reasoning in the control group is significantly (5%) larger compared to the proportion of Bayesian reasoning in the treatment group. These results prove that the use of Bayesian reasoning in Bayesian tasks decreases under time pressure.

Thirdly, heuristics are identified in only a small part of the sample. In total 19.38 percent of the subject was classified in one of the four heuristic categories. The last hypothesis stated that the use of heuristics in Bayesian tasks increases under time pressure. In the treatment group the use of heuristics is identified 17 times, in the control group only eight times. The proportions of the two groups are compared with Fisher's exact test. The test proves that the proportion of heuristics is significantly (5%) bigger in the treatment group compared to the control group. The use of heuristics in Bayesian tasks increases under time pressure.

Furthermore, a substantial number of participants (31.78%) calculated the joint probability instead of the posterior probabilities. This incorrect statistical rule is the most identified behavioural pattern in the control (39.39%) and treatment group (23.81%). To conclude, there is an outstanding part of the sample who used a random strategy. There is a random pattern found in 40.31 percent of the cases. As expected, with time pressure more people give random answers (46.03%) compared to the situation with no time pressure (34.85%). In Figure 4.5 the most frequent values given as an answer in each task are graphically displayed. For each task, the four highest peaks are identified as one of the behavioural patterns.

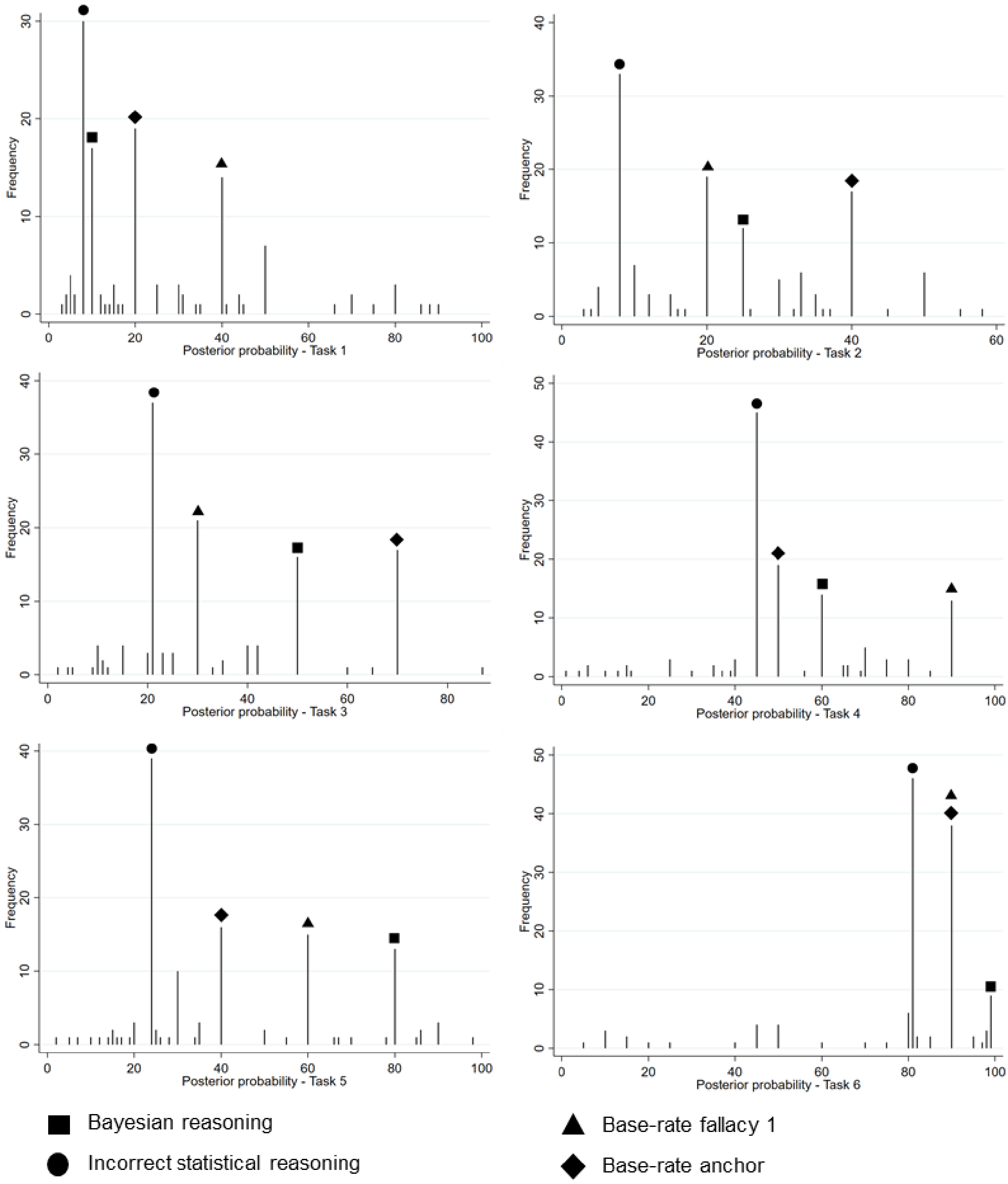


Figure 4.5: Spike-plots representing the frequency of answers given in each Bayesian task, peaks classified as behavioural patterns.

4.3 Demographic statistics

The demographic statistics give an indication of the robustness of the sample and the generalizability of the results. Both genders are well represented in the sample; 86 males and 43 females. The age of the participants is relatively low, 116 (90%) of the participants is in the age category 18-30. This unbalanced age distribution will expectedly not directly influence the results. The majority of the participants is highly educated; 129 (84%) participants finished a bachelor's degree, master's degree, PhD or a study at a University of Applied Sciences. Moreover, a part of the 16 participants who only finish secondary school is young and possibly following a study at a university at this moment. In line with this highly educated population, most people (74%) followed a statistical or mathematical course after secondary school. This information indicates the sample has on average a high cognitive ability and statistical/mathematical experience. There is no significant (10%) difference in the distribution of demographic characteristics between treatment and control group.

5 | Discussion and conclusion

The aim of this study is to clarify how and why time pressure affect peoples' Bayesian reasoning performance. It was hypothesised people act in line with Bayes' theorem when answering a Bayesian task with no time pressure. However, only 13.64 percent of the participants act in line with Bayes' theorem. The proportion of Bayesian subject is significantly lower than 50 percent. This does not affect the research goal; time pressure can still make decision makers be less precise in calculations and trigger changes in their behaviour. Which should cause a decrease in Bayesian reasoning performance. Although there is a small absolute difference in performance, the predicted decrease in Bayesian reasoning performance is not significant. Time pressure did not significantly reduce the quality of decision-making in Bayesian reasoning performance, based on analyses which control for cognitive ability, motivation, and experience. The Bayesian reasoning performance does not decrease but people's behaviour does change under time pressure. With no time pressure 13.64 percent of the people use Bayesian reasoning, under time pressure only 3.17 percent. This proportion is significantly smaller. On the other hand, the use of heuristics in Bayesian tasks increases under time pressure. With no time pressure 23.81 percent of the people use heuristics, under time pressure 39.39 percent. This proportion is significantly bigger. To conclude, time pressure does not decrease peoples' Bayesian reasoning performance significantly. However, time pressure trigger decision makers change their behaviour in Bayesian tasks. The use of Bayesian reasoning decreases and the use of heuristics increases. People make decisions more quickly, shift from a slow and optimal decision process to a faster and more autonomous process.

The first finding of this research is the surprisingly low number of people who act in line with Bayes' theorem with no time pressure. In studies with a similar research design, Bayesian reasoning performances were much better (Holt & Smith, 2009; El-Gamal & Grether, 1995). El-Gamal and Grether (1995) identified Bayesian reasoning as the most prominent rule used by their subjects. Our findings are in line with research by Kahneman and Tversky (1972), who claimed that people are: “not Bayesian et all.” (p. 450). A consequence of the low number of Bayesian subjects is the high deviation from the optimal posterior probability.

Why do people not act in line with Bayes’ rule? This research helps to exclude several factors. Subjects were on average motivated, intelligent and to some extent statistically educated. The fact that only one-third of the subjects were familiar with comparable tasks indicates ignorance as a possible cause for the observed performance. When a decision maker really does not know how to solve a Bayesian task, it does not matter whether you are intelligent, motivated and have the availability of unlimited time. The large number of subjects that use an incorrect statistical rule is in line with this reasoning, possibly people do think this is the correct way to solve the task. The use of an incorrect statistical rule is not mentioned in most researches, our research classified a large percentage of the subjects in this category. Did other research on Bayesian reasoning miss this pattern or did it simply not appear in their samples? In my opinion, previous research pays too little attention to the possibility that people do not know how to use Bayes’ rule. Most literature presumes it is a decision maker’s choice to deviate from the optimum. However, the results in this research indicate that people want to solve the task correctly but do not know how. In that case, it is not a choice to use a different behavioural pattern, it is a necessity.

Another explanation for the low number of Bayesian subject is misinterpretation of the task, caused by negligence by the subject or instructor. Subjects possibly thought they had to calculate the joint probability instead of the conditional probability. However, this is unlikely since the design of the experiment is similar to the design in previous research (El-Gamal & Grether, 1995; Grether, 1980). High motivation and intelligence decrease the probability of misinterpretation. Also, it is possible the subjects were not as motivated as they say they were. The monetary incentive is relatively small, peoples’ motivation is mostly intrinsic driven. Is this motivation high in off to invest the time and mental effort needed in the tasks? The fact that subjects report their motivation after completing the Bayesian tasks argues in favour of an actual high motivation. The significant decrease in performance between the first and last three tasks, both in the control and treatment group, can indicate a lack of motivation. However, this difference is mainly caused by the performance in the fifth task. People in both groups performed worst in this task, possibly this task was more difficult.

There is a clear effect of time pressure, the number of Bayesian subjects decreases significantly. Time pressure increases the level of stress, pressure, and arousal, this affects people decision-making processes (Maule & Hockey, 1993). People do not have or think they do not have enough time to calculate the Bayesian optimum. Time pressure increases the use of heuristic processes; especially when people have high motivation to process information (Suri & Monroe, 2003). These efficient cognitive processes lower time and metal costs (Gigerenzer & Gaissmaier, 2011; Shah & Oppenheimer, 2008), but lead to systematic errors or biases (Tversky & Kahneman, 1974). Time pressure triggers decision makers to deviate from the economic optimal decision.

Most surprisingly, time pressure does not cause a significant decrease in Bayesian reasoning performance. Although, the shift from Bayesian reasoning to heuristic does cause a decrease in Bayesian reasoning performance. Due to the small number of Bayesian subjects, only a small part of the sample made this shift. Moreover, the average Bayesian reasoning performance in the random category is comparable to the performance of people who used an incorrect statistical rule. The higher number of random participants in the treatment group outweigh the number of people who used an incorrect statistical rule in the control group. Important to note, the results do not prove that the use of heuristics is as effective as Bayesian reasoning.

This paper has limitations, mostly caused by time and financial stints. Most importantly, a larger sample would be valuable. This will give the statistical test more power and small observed differences between the groups can be significant when observed in larger samples. Secondly, it could be interesting to execute the experiment with a higher monetary reward. The probability to earn money in this experiment is low, in my opinion too low to truly motivate participants. There is some evidence that an increase in financial motivation results in more Bayesian subject (Grether, 1980). Besides, higher incentives make it possible to ask the participants to perform more tasks which will help to identify behavioural patterns. Ideally, this future experiment will be executed in the lab. This controlled environment helps to exclude factors which can influence the results, such as distraction or tools. Thirdly, I would advise testing extra elements which will help to clarify observed behaviour. It is interesting to see whether participants think they gave the correct answer. This can help to clarify whether the deviation from the optimal decision is a conscious or unconscious decision. Also, it can be valuable to test whether the participant understood the task. This can exclude misunderstanding as a possible explanation of economic non-optimal behaviour. Next, measuring motivation before and after the tasks, can help to explain the difference in motivation between the control and treatment group. Is this difference caused by the time pressure? Furthermore, future research with variation in the type of Bayesian tasks, variation in time limitation and variation in the reward system can be valuable.

This research contributes to a better understanding of individual decision processes, specifically when facing a Bayesian task. The thesis helps to explain how, why and when decision makers actions differ from Bayes' theorem. The research proved that time pressure triggers people to use heuristics instead of Bayesian reasoning, observed the rareness of Bayesian reasoning and the frequent use of incorrect statistical reasoning, both with and without time pressure. Understanding these biases can help to design better mechanisms to predict behaviour and a better understanding of decision-making under uncertainty. The knowledge on heuristics can be implemented in the consumer products market, policy areas, insurances, voting or medical considerations (Grether, 1980). Time constraints play a role in nearly all decisions in daily life, to a greater or lesser extent. By understanding people's economic behaviour under time pressure, it is possible to inform, help or manipulate people more efficiently. To conclude, it is remarkable that intelligent, motivated and statistical sophisticated subjects are not able to solve Bayesian tasks correctly. This is not very hopeful for less motivated or intelligent decision makers. I would like to recommend special attention for statistical education in the fields in which Bayesian reasoning is applied commonly. In other fields, it is better to avoid Bayesian reasoning. Choose other ways to communicate this kind of information, especially in situations with time pressure.

References

Achtziger, A., & Alós-Ferrer, C. (2013). Fast or rational? A response-times study of Bayesian updating. *Management Science*, 60(4), 923-938.

Ahituv, N., Igbaria, M., & Sella, A. V. (1998). The effect of time pressure and completeness of information on decision making. *Journal of management information systems*, 15(2), 153-172.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly Harmless Econometrics.* Princeton and Oxford, Chapter 2: Princeton University Press .

Ariely, D., Loewenstein, G., & Prelec, D. (2003). Coherent arbitrariness: Stable demand curves without stable preferences. *The Qualterly journal of economics*, 118(1), 73-106.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44(3), 211-233.

Bar-Hillel, M. (1983). The base rate fallacy controversy. *Advances in Psychology*, 16, 39-61.

Bayes, T. (1764). An Essay Toward Solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London*, 53, 370-418.

Ben Zur, H., & Breznitz, S. J. (1981). The effect of time pressure on risky choice behavior. *Acta Psychologica*, 47(2), 89-104.

Benjamin, D. J. (2018). *Errors in probablistic reasoning and judgment biases* . Cambridge: National Bureau of Economic Research .

Benson, L., & Beach, L. R. (1996). The effect of time constraints on the prechoice screening of decision options. *Organizational Behavior and Human Decision Processes*, 67(2), 222-228.

Bergman, O., Ellingsen, T., Johannesson, M., & Svensson, C. (2010). Anchoring an cognitive ability. *Economics Letters*, 66-68.

Bobadilla-Suarez, S., & Love, B. C. (2018). Fast or frugal, but not both; Decision heuristics under time pressure. *Journal of Experimental Psychology: learning, Memory, and Cognition*, 44(1), 24.

Bogasz, R., Wagenmakers, E. J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in neurosciences,* 33(1), 10-16.

Brase, G. L., & Hill, W. T. (2015). Good fences make for good neighbors but bad science: a review of what improves Bayesian reasoning and why. *Frontiers in Psychology* , 6, 340.

Camerer, C. F., & Hogarth, R. M. (1999). The effect of financial incentives in experiments: A review and capital-labor-production framework. *Journal of risk and uncertainty*, 19(1-3), 7-42.

Casscells, W., Schoenberger, A., & Graboys, T. B. (1978). Interpretation by physicians of clinic laboratory results. *New England Journal of medicine*, 299(18), 999-1001.

Chapman, G. B., & Johnson, E. J. (2002). Incorporating the irrelevant: Anchors in judgements of belief and value. *Heuristics and biases: The psychology of intuitive judgment*, 120-138.

Christensen-Szalanski, J. J., & Beach, L. R. (1982). Experience and the base-rate fallacy. *Organizational Behavior and Human Performance*, 29(2), 270-278.

Compte, O., & Postlewaite, A. (2018). *Ignorance and Uncertainty*. Cambridge: Cambridge University Press.

Cox, C., & Mouw, J. T. (1992). Disruption of the representativeness heuristic: Can we be perturbed into using correct probabilistic reasoning? *Educational Studies in Mathematics,* 23(2), 163-178.

Croson, R., & Gächter, S. (2010). The science of experimental economics. *Journal of Economic Behavior & Organization*, 73(1), 122-131.

Dhar, R., & Nowlis, S. M. (1999). The effect of time pressure on consumer choice deferral. *Journal of Consumer Research*, 25(4), 169-384.

Dinner, I., Johnson, E. J., Goldstein, D. G., & Liu, K. (2011). Partitioning default effects: why people choose not to choose. *Journal of Experimental Psychology: Applied*, 17(4), 332-341.

Donkin, C., Little, D. R., & Houpt, J. W. (2014). Assessing the speed-accuracy trade-off effect on the capacity of information processing. *Journal of Experimental Psychology: Human Perceptiono and Performance*, 40(3), 1183-1202.

Eddy, D. M. (1982). Probalistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky, *Judgment under uncertainty: Heuristics and biases* (pp. 249-267). Cambridge, England: Cambridge University Press.

Edland, A., & Svenson, O. (1993). Judgment and decision making under time pressure. In O. Svenson, & A. Maule, *Time pressure and stress in human judgment and decision making* (pp. 27-40). New York: Plenum press .

El-Gamal, M. A., & Grether, D. M. (1995). Are people Bayesian? Uncovering behavioral strategies. *Journal of the American statistical Association*, 90(432), 1137-1145.

Enough, B., & Mussweiler, T. (2001). Sentencing Under Uncertainty: Anchoring effects in the courtroom 1. *Journal of applied social psychology,* 31(7), 1535-1551.

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differntial processing of self-generated and experimenter-provided anchors. *Psychological science*, 12(5), 391-396.

Epley, N., & Gilovich, T. (2005). When effortful thinking influences judgmental anchoring: differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making*, 18(3), 199-212.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient . *Psychological science*, 17(4), 311-318.

Evans, J. S., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3), 223-241.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4), 25-42.

Frederiks, E. R., Stenner, K., & Hobman, E. V. (2015). Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour. *Renewable and Sustainable Energy Reviews*, 41, 1385-1394.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The journal of socio-ecnomics*, 40(1), 35-42.

Furnham, A., & Chamorro-Premuzic, T. (2004). Personality and intelligence as predictors of statstical examinatino grades. *Personality and individual differences*, 37(5), 943-955.

Gehlbach, H., & Barge, S. (2012). Anchoring and adjusting in questionnaire responses. *Basic and Applied Social Psychology*, 34(5), 417-433.

Gigerenzer. (2008). Why heuristics work. *Perspectives on psychological science*, 3(1), 20-29.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, 62, 451-482.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In *Simple heuristics that make us smart* (pp. 3-34). Oxford University Press.

Goodie, A. S., & Crooks, C. L. (2004). Time-pressure effects on performance in a base-rate task. *The journal of general psychology*, 131(1), 18-28.

Grether, D. M. (1980). Bayes rule as a descriptive model: The representativeness heuristic. *The Quarterly Journal of Economics*, 17(1), 31-57.

Grether, D. M. (1992). Testing Bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior & Organization*, 17(1), 31-57.

Hahn, M., Lawson, R., & Lee, Y. G. (1992). The effect of time pressure and information load on decision quality. *Psychology & Marketing*, 9(5), 365-378.

Haigh, M. (2016). Has the standard cognitive reflection test become a victim of its own succes? *Advances in Cognitive Psychology*, 12(3), 145-149.

Harrison, G. W., Martínez-Correa, J., & Swarthout, J. T. (2014). Electing subjective probabilities with binary lotteries. *Journal of Economic Behavior & Organization*, 101, 128-140.

Holt, C. A., & Smith, A. M. (2009). An update on Bayesian updating. *Journal of Economic Behavior & Organization*, 69(2), 125-134.

Huerta, M. P., & Lonjedo, M. A. (2007). Proceedings of the Fifth Congress of the European Society for Research in Mathemtics Education. *The same problem in three presentation formats: Different percentages of success and thinking processes* (pp. 732-741). Cyprus: Department of Education - University Cyprus.

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, 1161-1166.

Joyce, E. J., & Biddle, G. C. (1981). Anchoring and adjustment in probabilisitc inference autiting. *Journal of accounting* , 120-145.

Kahneman, D., & Frederick, S. (2002). Representativeness revisted: Attribute substitution in intuitive judgment. In *Heuristics of Intuitive Judgment: Extentions and Applications* (pp. 49-80). New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3(3), 430-454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, 80(4), 237.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challanges. *Behavioral and brain sciences*, 19(1), 1-17.

Levin, D., Peck, J., & Ivanov, A. (2016). Separating Bayesian updating from non-probalistic reasoning: An experimental investigation . *American Economic Journal: Microeconomics*, 8(2), 39-60.

Manrai, A. K., Bhatia, G., Strymish, J., Kohane, I. S., & Jain, S. H. (2014). Medicine's uncomfortable relationship with math: calculating positive predictive value . *JAMA internal medicine*, 174(6), 991-993.

Maule, A. J., & Hockey, G. R. (1993). State, stress, and time pressure . In *Time pressure and stress in human judgment and decision making* (pp. 83-101). Boston, MA: Springer.

Maule, A. J., Hockey, G. R., & Bdzola, L. (2000). Effect of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. *Acta psychologica*, 104(3), 283-301.

McDaniel, L. S. (1990). The effect of time pressure and audit program structure on audit performance. *Journal of Accounting Research*, 28(2), 267-285.

Morgan, D. (2018, October 5). What the test don't show. *Washington Post*, p. Retrieved from: www.washingtonpost.com.

Mousavi, S., & Gigerenzer, G. (2014). Risk, uncertainty, and heurstics. *Journal of Business Research*, 67(8), 1671-1678.

Ordonez, L., & Benson, L. (1997). Decisions under time pressure: How time constraint affects risky decision making. *Organizational Behavior and Human Decisioin Processes*, 71(2), 121-140.

Payne, J., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* Cambridge univeristy press.

Pearlin, L. I. (1961). The appeals of anonymity in questionnaire response. *Public Opinion Quarterly*, 25(4), 640-647.

Philips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of experimental psychology*, 72(3), 346-354.

Pollatsek, A., Well, A. D., Konold, C., Hardiman, P., & Cobb, G. (1987). Understanding conditional probabilities. *Organizational Behavior and Human Decision Processes*, 40(2), 255-269.

Read, D. (2005). Monetary incentives, what are they good for? *Journal of Economic Methodology*, 12(2), 265-276.

Salop, S. C. (1987). Salop, S. C. (1987). Evaluating uncertain evidence with Sir Thomas Bayes: A note for teachers. *Journal of Economic Perspectives*, 1(1), 155-159.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological bulletin*, 134(2), 207-222.

Smith, V. L. (1982). Microeconomic sytems as an experimental science. *The American Economic Review*, 72(50), 923-955.

Stieger, S., & Reips, U. D. (2016). A limitation of the cognitive reflection test: familiarity. *PeerJ*, e2395.

Suri, R., & Monroe, K. B. (2003). The effect of time constraints on consumers' judgments of price and products. *Journal of consumer research* , 30(1), 92-104.

Svenson, O., Edland, A., & Slovic, P. (1990). Choices and judgments of incompletely described decision alternatives under time pressure. *Acta Psychologica*, 75(2), 153-169.

Taffler, R. J. (2010). The representativness heuristic . *Behavioral finance: Investors, corporations, and markets*, 259-276.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), 207-232.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briére, N. M., Senecal, C., & Valliéres, É. F. (1993). On the assessment of intrinsic, extrinic, and amotivation in education: Evidence on the concurrent and construct validity of the Academic Motivation Scale. . *Education and psychological measurement*, 53(1), 159-172.

Weenig, M. W., & Maarleveld, M. (2002). The impact of time constraint on information search strategies in complex choice tasks. *Journal of Economic Psychology*, 23(6), 689-702.

Wegener, D. T., Petty, R. E., Blankenship, K. L., & Detweiler-Bedell, B. (2010). Elaboration and numerical anchoring: Implications of attitude theories for consumer judgment and decision making. *Journal of Consumer Psychology*, 20(1), 5-16.

Wells, G. L. (1985). The conjunction error and the representativeness heuristic. *Social Cognition*, 3(3), 266-279.

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, 125(4), 387.

Yelle, L. E. (1979). The learning curve: Historical review and comprehensive survey. *Decision sciences*, 10(2), 302-328.

Appendixes

Appendix A – Cognitive Reflection Test

1. *A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? _____ cents*

2. *If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes*

3. *In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days*

Appendix B – Experiment

Introduction

Erasmus School of Economics

This is an experiment for my master thesis at the Erasmus University Rotterdam.

The experiment consists of 15 questions and will take approximately 8 minutes of your time. Your answers will be kept strictly confidential and processed anonymously.

Your participation is greatly appreciated!

Sincerely,
Jan

In case of any questions, please contact me via email: 511715jg@student.eur.nl

→

Erasmus School of Economics

By filling in this survey you get the chance to win €30,-.

You will face 6 tasks, in each task you can earn lottery tickets. The more tickets you earn, the higher your chance to win the €30,-. The number of lottery tickets is linearly correlated with the correctness of your answers. You will earn 4 tickets for the correct answer. The higher the deviation from the correct answer, the lower number of tickets you will earn.

The better you answer, the higher your chance to win the lottery.

After two weeks, one winner will be selected by picking one of the lottery tickets with a random draw.

If you want to participate for the price, please fill in your email address at the end of the survey.

→

Bayesian tasks (treatment)

Erasmus School of Economics

Please, read this information carefully. The next 6 tasks will have the following design:

Suppose there are three boxes; box X, box A, box B. Inside box X are 10 balls, all marked with an ‘A’ or a ‘B’. Inside box A and B are 10 balls, all red or blue. The table below shows an example of the type and number of balls in each box.

Box X
1 'A' & 9 'B'

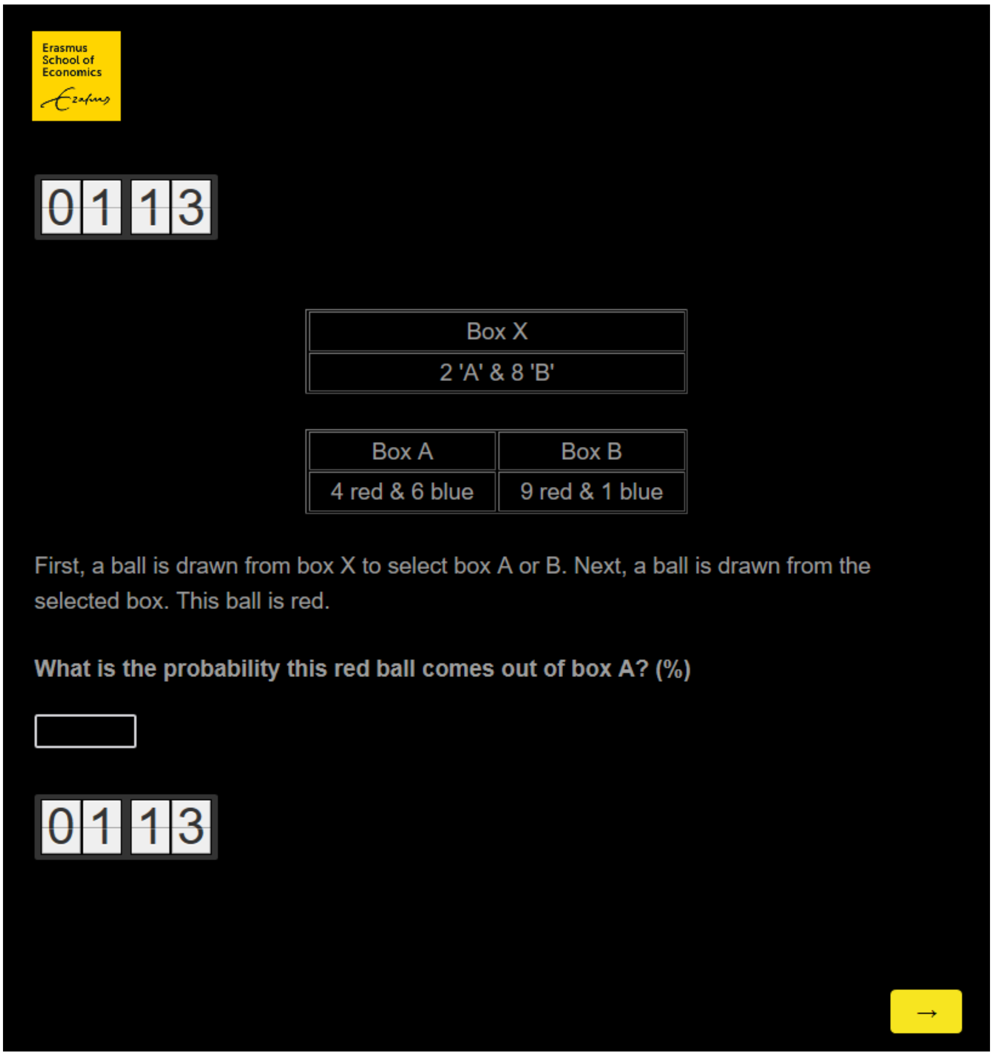
Box A	Box B
6 red & 4 blue	3 red & 7 blue

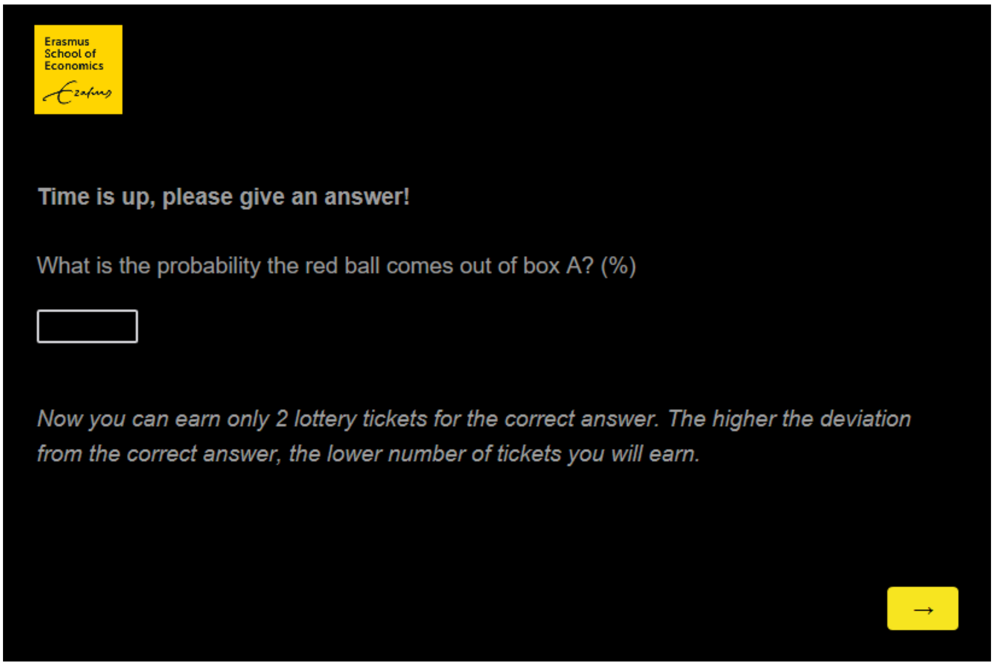
First, a ball is drawn from box X. The letter on this ball determines from which box the next draw will be made. However, you do not know which ball is selected from box X.

Next, a ball is drawn from the selected box (box A or box B). This ball will be shown to you.

Given the colour of this ball, it is your task to provide the probability this ball comes out of box A.

You will have limited time for each task, a timer will indicate how many seconds you have left to give an answer!


Erasmus School of Economics
01 13
Box X
2 'A' & 8 'B'
Box A
Box B
4 red & 6 blue
9 red & 1 blue
First, a ball is drawn from box X to select box A or B. Next, a ball is drawn from the selected box. This ball is red.
What is the probability this red ball comes out of box A? (%)
01 13
→


Erasmus School of Economics
Time is up, please give an answer!
What is the probability the red ball comes out of box A? (%)
Now you can earn only 2 lottery tickets for the correct answer. The higher the deviation from the correct answer, the lower number of tickets you will earn.
→

Bayesian task (control)

Please, read this information carefully. The next 6 tasks will have the following design:

Suppose there are three boxes; box X, box A, box B. Inside box X are 10 balls, all marked with an 'A' or a 'B'. Inside box A and B are 10 balls, all red or blue. The table below shows an example of the type and number of balls in each box.

Box X
1 'A' & 9 'B'

Box A	Box B
6 red & 4 blue	3 red & 7 blue

First, a ball is drawn from box X. The letter on this ball determines from which box the next draw will be made. However, you do not know which ball is selected from box X.

Next, a ball is drawn from the selected box (box A or box B). This ball will be shown to you.

Given the colour of this ball, it is your task to provide the probability this ball comes out of box A.

Box X
2 'A' & 8 'B'

Box A	Box B
4 red & 6 blue	9 red & 1 blue

First, a ball is drawn from box X to select box A or B. Next, a ball is drawn from the selected box. This ball is red.

What is the probability this red ball comes out of box A? (%)

→

Questionnaire

I was motivated to answer the tasks correctly

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Are you familiair with similar tasks?

- Yes
- No

Erasmus School of Economics

A bat and a ball cost $ 1.10 in total. The bat costs $ 1.00 more than the ball.
How much does the ball cost? (cents)

If it takes 5 machines 5 minutes to make 5 widgets.
How long would it take 100 machines to make 100 widgets? (minutes)

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake.
How long would it take for the patch to cover half of the lake? (days)

→

Erasmus School of Economics

What is your gender?

- Male
- Female
- Other

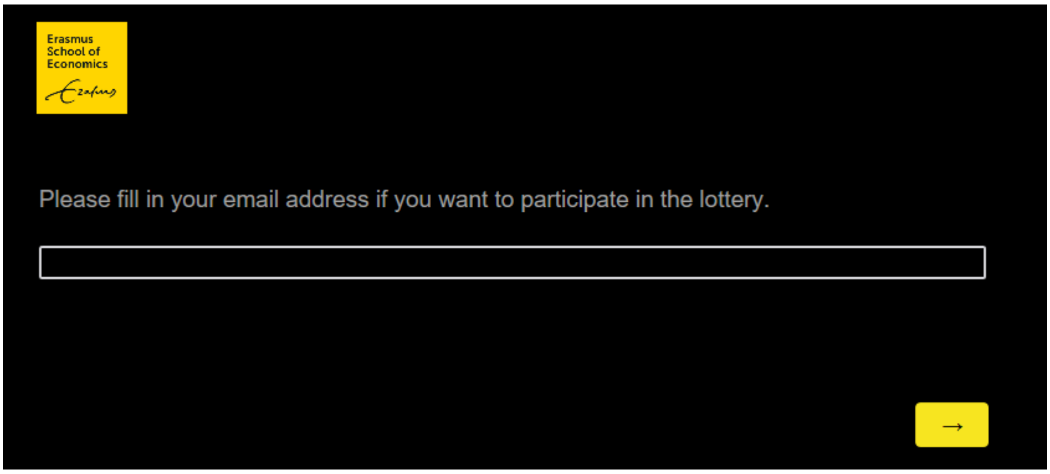
What is your age?

What is the highest level of education you have completed?

- Primary school
- Secondary school
- Vocational education
- University of Applied Sciences
- Bachelor's degree (University)
- Master's degree (University)
- PhD

Have you learned any statistical or mathematical courses after secondary school?

- Yes
- No


Erasmus School of Economics
Please fill in your email address if you want to participate in the lottery.
→

Appendix C – Test results

Variable	Variable	Test	Test-value
Group (treatment/control)	Gender	1-tailed Fisher's exact	0.2880
Group (treatment/control)	Age	2-tailed Mann-Whitney U	0.1817
Group (treatment/control)	Age (categories)	2-tailed Mann-Whitney U	0.1744
Group (treatment/control)	Education	Fisher's exact	0.5790
Group (treatment/control)	Statistical course	1-tailed Fisher's exact	0.1230
Group (treatment/control)	Cognitive ability	2-tailed Mann-Whitney U	0.1321
Group (treatment/control)	Motivation	2-tailed Mann-Whitney U	0.0383
Group (treatment/control)	Motivation	1-tailed Mann-Whitney U	0.0193
Group (treatment/control)	Experience	1-tailed Fisher's exact	0.3470
Number Bayesian subject (control)	>9	Binominal	0.0001
Number Bayesian subject (control)	≥50%	One-sample proportion	0.0001
Group (treatment/control)	Bayesian subjects	1-tailed Fisher's exact	0.0330
Group (treatment/control)	Heuristic subjects	1-tailed Fisher's exact	0.0270
Group (treatment/control)	BRP* task 1-6	2-tailed Mann-Whitney U	0.5354
Group (treatment/control)	BRP task 1-6	1-tailed Mann-Whitney U	0.2685
Group (treatment/control)	BRP task 1-3	2-tailed Mann-Whitney U	0.9046
Group (treatment/control)	BRP task 1-3	1-tailed Mann-Whitney U	0.7426
Group (treatment/control)	BRP task 4-6	2-tailed Mann-Whitney U	0.5164
Group (treatment/control)	BRP task 4-6	1-tailed Mann-Whitney U	0.5487
Group (treatment/control)	BRP task 1	2-tailed Mann-Whitney U	0.0264
Group (treatment/control)	BRP task 2	2-tailed Mann-Whitney U	0.8519
Group (treatment/control)	BRP task 3	2-tailed Mann-Whitney U	0.9904
Group (treatment/control)	BRP task 4	2-tailed Mann-Whitney U	0.1044
Group (treatment/control)	BRP task 5	2-tailed Mann-Whitney U	0.5589
Group (treatment/control)	BRP task 6	2-tailed Mann-Whitney U	0.7730
Group (treatment/control)	Time used task 1-3	2-tailed Mann-Whitney U	0.0005
Group (treatment/control)	Time used task 1-3	1-tailed Mann-Whitney U	0.0210
Group (treatment/control)	Time used task 4-6	2-tailed Mann-Whitney U	0.0044
Group (treatment/control)	Time used task 4-6	1-tailed Mann-Whitney U	0.1046
BRP task 1-3 (treatment)	BRP task 4-6 (treatment)	2-tailed Wilcoxon	0.0001
BRP task 1-3 (treatment)	BRP task 4-6 (treatment)	1-tailed Wilcoxon	0.0001
BRP task 1-3 (control)	BRP task 4-6 (control)	2-tailed Wilcoxon	0.0001
BRP task 1-3 (control)	BRP task 4-6 (control)	1-tailed Wilcoxon	0.0001

**Bayesian reasoning performance (BRP)*

Appendix D – Behavioural patterns

	Base-rate	Diagnostic information	
#	*Box X (a/b)*	*Box A (r/b)*	*Box B (r/b)*
1	2/8	4/6	9/1
2	4/6	2/8	4/6
3	7/3	3/7	7/3
4	5/5	9/1	6/4
5	4/6	6/4	1/9
6	9/1	9/1	1/9

#	1	2	3	4	5	6
Bayesian reasoning	*10*	*25*	*50*	*60*	*80*	*99*
Base-rate anchor	*20*	*40*	*70*	*50*	*40*	*90*
Base-rate fallacy 1	*40*	*20*	*30*	*90*	*60*	*90*
Base-rate fallacy 2	*31*	*33*	*30*	*60*	*86*	*90*
Incorrect statistical rule	*8*	*8*	*21*	*45*	*24*	*81*

Bayesian reasoning **:** *Correct use of Bayes' rule.*

*Task 1 : (0.2*0.4)/((0.2*0.4)+(0.8*0.9)) = 10*
*Task 2 : (0.4*0.2)/((0.4*0.2)+(0.6*0.4)) = 25*
*Task 3 : (0.7*0.3)/((0.7*0.3)+(0.3*0.7)) = 50*
*Task 4 : (0.5*0.9)/((0.5*0.9)+(0.5*0.6)) = 60*
*Task 5 : (0.4*0.6)/((0.4*0.6)+(0.6*0.1)) = 80*
*Task 6 : (0.9*0.9)/((0.9*0.9)+(0.1*0.9)) = 99*

Base-rate anchor **:** *Use the base-rate as an anchor, no or insufficient use of diagnostic information.*

Base-rate fallacy 1 **:** *Answer based on the percentage of red balls in box A, no or insufficient use of base-rate information.*

Base-rate fallacy 2 **:** *Answer based on the percentage of red balls in box A out of the total number of red balls. No or insufficient use of base-rate information.*

Task 1 : 0.4/(0.4+0.9) = 31
Task 2 : 0.2/(0.2+0.4) = 33
Task 3 : 0.3/(0.3+0.7) = 30
Task 4 : 0.9/(0.9+0.6) = 60
Task 5 : 0.6/(0.6+0.1) = 86
Task 6 : 0.9/(0.9+0.1) = 90

Inc. statistical reasoning **:** *Calculation based on probability red and box A not considering the colour of the ball is already known.*

Task 1	:	*(0.2*0.4)*	=	*8*
Task 2	:	*(0.4*0.2)*	=	*8*
Task 3	:	*(0.7*0.3)*	=	*21*
Task 4	:	*(0.5*0.9)*	=	*45*
Task 5	:	*(0.4*0.6)*	=	*24*
Task 6	:	*(0.9*0.9)*	=	*81*