To eat humble pie – the effect of the Bayesian Truth Serum on food choices

An experimental approach

A Master's thesis by

F.J.M. de Rooij

Student number: 385785fr

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

MSc Economics & Business

Specialisation Behavioural Economics

Thesis supervisor: J.T.R. Stoop

Second assessor: S.C. van der Zee

Date: 29 August 2018

PREFACE AND ACKNOWLEDGEMENTS

During my bachelor Economics and Business Economics, the field of Behavioural Economics immediately aroused my interest. This made the choice to apply for the master Behavioural Economics easy. As I already followed the Financial Economics track in my bachelor, I chose to follow the same track in my master. I followed the corresponding courses with pleasure, but the other Behavioural Economics courses also aroused my curiosity about the other areas to which Behavioural Economics can be applied. Therefore, I chose to write my thesis not about a Financial Economics topic. I wanted to write about a subject that people face on a daily basis: food. For many people the relationship with food is not always that healthy, which has resulted in many dieting trends. Moreover, there is still a taboo on having a certain weight or eating pattern. Combined with my interest in lying behaviour, this resulted in the following thesis topic: “The effect of the Bayesian Truth Serum on food choices”. It seemed interesting to me to investigate whether ‘eating humble pie’ affects someone’s eating behaviour. The result lies in front of you.

This thesis would not have been possible without the help and support of several people. First, I would like to thank my supervisor J.T.R. Stoop for his support and useful feedback. I very much appreciate the fact that even during his holiday he took the time to read my documents, and that he was able to teach me something about healthy food as well. I would also like to thank S.C. van der Zee to be my second reader. Finally, I would like to thank my family and friends for their support during the writing process and for occasionally taking care of the necessary distractions.

Fréderique de Rooij

Rotterdam, August 2018

ABSTRACT

Over the last few years obesity has become a worldwide health and healthcare issue. Food tracking might help to solve this problem for some people, but the effect of self-monitoring methods can be ambiguous due to several behavioural factors. The behavioural factor of interest in this study is being (dis)honest about eating behaviour, which can be caused by social desirability bias. This thesis examines whether people make healthier choices if they are incentivized to be honest about their eating behaviour, by using the Bayesian Truth Serum. An electronic questionnaire is used to elicit current self-reported food intakes, and to let the respondents make hypothetical meal choices. About half of the respondents only had to report their own answers, whereas the other half had to answer the questions according to the Bayesian Truth Serum method. No evidence is found that the respondents made healthier food choices if they were exposed to the Bayesian Truth Serum, nor that they reported their current intakes more truthfully. Thus, incentivizing honest behaviour does not generate healthier decisions. These findings indicate that the Bayesian Truth Serum is not an effective measure to improve food-tracking methods to help solve the obesity issue.

Keywords: *Bayesian Truth Serum, hypothetical meal choices, self-report measures, Social Desirability Bias, eating behaviour*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF GRAPHS

APPENDICES

1. INTRODUCTION

With the rise of food blogs, cookbooks and diets, 'healthy' food and diet trends arose (Pel, 2016). Nowadays, people are more conscious about what they eat and how this influences their image (RTL Nieuws, 2016). Contrarily, people have less time to cook, making fast food more appealing to eat (Carlak, 2018). Consequently, obesity has become a worldwide problem and 48.7 per cent of the Dutch people were overweight in 2017 (Centraal Bureau voor de Statistiek, 2018). In 2010, the costs of obesity were approximately 3.2 billion euro for the Dutch society (ANP, 2010).

As a result, many people try to lose weight, but most attempts remain unsuccessful. People overestimate their willpower and cannot comply with the strict rules that they impose on themselves (Van der Zee, 2016). Moreover, most people are inaccurate in estimating their body size and often lack taking actions for weight control (Blokstra, Burns, & Seidell, 1999).

Something that might work in the fight against overweight, is self-monitoring, i.e. keeping track of eating and exercise behaviour by using activity trackers or by keeping a food diary (Burke, Wang, & Sevick, 2011). Research shows that using an activity tracker increases the physical activity of obese adults (de Vries, Kooiman, van Ittersum, van Brussel, & de Groot, 2016). Activity trackers are less susceptible to the thoughts and answers of subjects than food diaries are as they usually keep track of activity automatically.

The effect of food tracking can be ambiguous. If used properly and truthfully, it could give insights which help changing eating behaviour (Burke, Wang, & Sevick, 2011). However, the success rate of food tracking depends on several behavioural factors: the promptness, the consistency and the truthfulness of self-monitoring (Bandura, 1998). The focus of this thesis will be on the latter: (dis)honesty of people about their eating behaviour.

Previous research in the United States has shown that 31 per cent of the participants underreported their dietary intake (Klesges, Eck, & Ray, 1995). Especially obese persons underreported their calorie intake, varying from 50 to 80 per cent of their measured energy expenditure (Lichtman et al., 1992; Bandini, Schoeller, Cyr, & Dietz, 1990). In order to mitigate the discrepancy between the self-reported and actual calorie intake, one ought to know whether underreporting is done intentionally or unconsciously. Muhlheim, Allison, Heshka, and Heymsfield (1998) have examined whether unsuccessful dieters intentionally underreport their food intake. The intentions of the participants were elicited by using a bogus pipeline paradigm, i.e. deceiving the subjects to get truthful answers. The belief that their self-reported calorie intake could be validated by the researchers improved the accuracy of the self-reported intake, but did not restrain the participants from underreporting.

Underreporting can be caused by the tendency to give socially desirable responses (Paulhus, 1991). Socially desirable responding can be induced by internal and external factors, i.e. self-deception and impression management respectively (Gur & Sackheim, 1979; Paulhus, 1986).

One way to reduce self-deception in self-report questionnaires is the Bayesian Truth Serum (Bernardic, 2017). The Bayesian Truth Serum is a survey scoring method that can be used to elicit truthful subjective information (Prelec, 2004). Instead of assigning high scores to the most common answers, the answers that are more common than collectively predicted get high scores. Even if someone has a minority opinion, one can still maximise his expected score. Thus, this should restrain participants from giving answers that they think is most common, i.e. closest to the mean of the population. In other words, participants will be incentivized to give honest answers.

It would be interesting to examine whether honest answers about someone's eating behaviour can be elicited without deceiving the participants and whether being more honest induces healthier choices. This leads to the following research question:

Do people make healthier choices if they are incentivized to be honest about their eating behaviour?

In the following chapter, the theoretical framework will be outlined. Socially desirability bias and its two components, self-deception and impression management, will be discussed. Moreover, the Bayesian Truth Serum method will be thoroughly explained.

In chapter three, a description of the questionnaire that is used to gather data is given. All participants will have to answer a survey about their eating behaviour, while the treatment group will also have to answer the questions according to the Bayesian Truth Serum. Moreover, the gathered dataset will be described.

Subsequently, the methodology that is used to analyse the data will be illustrated in chapter four. The results of this analysis will be discussed in chapter five.

Next, limitations of this research and recommendations for further research will be presented in the discussion.

Finally, the research question will be answered in the conclusion.

2. THEORETICAL FRAMEWORK

The introduction described the general motivation behind the research question. More specifically, research shows that people are not honest about their dietary intake in self-report questions. Applying the Bayesian Truth Serum could incentivize people to be more honest. However, the effect of the Bayesian Truth Serum on reporting food intake has not been studied before.

In this chapter, the theory behind the motivation for the research question will be discussed.

The first part focuses on providing a possible reason for dishonesty regarding self-monitoring of dietary intake: the social desirability bias. Social desirability bias can be divided into two components: self-deception and impression management. As the focus of this thesis will be on the effect of the Bayesian Truth Serum on food choices, the level of social desirability bias (and its two components) will not be measured. However, the literature will show that the Bayesian Truth Serum decreases the level of social desirability bias, which should affect the answers of the subjects.

The second part of this chapter will be focussed on the theory behind the methodology that is used to limit self-deception, called the Bayesian Truth Serum.

Finally, the hypotheses that will be tested are described.

2.1 Social Desirability Bias

Social desirability bias (SDB) can be described as the tendency of humans to answer questions in a manner that one presents the best possible version of oneself (Paulhus, 1991). Paulhus, Harms, Bruce, and Lysy (2003) found that people generally overstate the positivity of their self-reported answers. As researchers often employ self-report measures in questionnaires and rely on these questions to be answered truthfully, SDB can distort the information gained from these self-reports (Fisher, 1993). This makes SDB a prevalent cause for comprising the validity of experimental findings in psychology research (Nederhof, 1985; Paulhus, 1991). More specifically, Nederhof (1985) found that 10 to 75 per cent of the variance in the responses in a study can be explained by SDB. Consequently, the conclusions drawn from such a study could point in another direction than what should be concluded based on actual behaviour. The probability that socially desirable responses are given increases with the level of social sensitivity of the respective question (King & Bruner, 2000).

Participants can be either unwilling or unable to answer these questions precisely, which results in systematically biased data toward the participants' perceptions of what is socially desirable (Maccoby & Maccoby, 1954).

There are several reasons why participants do not respond truthfully to self-report questionnaires. First, this behaviour can be the consequence of the desire to confirm to social norms, avoid criticism, protect the self, or acquire social approval (Huang, Liao, & Chang, 1998). Secondly, the maintenance of both personal mental health and interpersonal relationships can be seen as justification for the deception of self and others (Taylor & Brown, 1988). Moreover, people incorporate the norms and values of their community as an internal reference point against which someone compares his own actions (Henrich, et al., 2001). Socially desirable responding can be used to avoid the internal and external punishment caused by noncompliance with the internal benchmark (Mazar, Amir, & Ariely, 2008).

Applied to a dietary context, self-deceptive behaviour can protect dieters from condemnation by a society that uses thin people as a reference point and considers overweight to be one's own fault (Crandall, 1994; Muhlheim et al., 1998).

Previous research has divided SDB into two components, depending on the perceived audience: self-deception and other-deception, which is also known as impression management (Sackheim & Gur, 1979; Paulhus, 1986). Self-deception is solely intended for the self, while the latter is concentrated on the image shown to others. These two terms will be discussed more thoroughly in the following two subparagraphs.

2.1.1 Self-deception

The first dimension of SDB refers to an internal process: self-deception. Self-deceptive behaviour entails that people distort information primarily for oneself (Roth, Snyder, & Pace, 1986). Thus, people actually believe that their positive answers in self-report questionnaires are accurate (Paulhus, 1984). Even though individuals know the truth deep-inside, they decide to believe something contradictory.

Sartre (1958) described this paradox as follows: *"It follows first that the one to whom the lie is told and the one who lies are one and the same person, which means that I must know in my capacity as deceiver the truth which is hidden from me in my capacity as the one deceived. Better yet I must know the truth very exactly in order to conceal it more carefully – and this not at two different moments, which at a pinch would allow us to re-establish a semblance of duality – but in the unitary structure of a single project. How then can the lie subsist if the duality which conditions it is suppressed?"*

Sackeim and Gur (1979) resolved this paradox by presenting four criteria that one should met to be regarded as self-deceptive. First, an individual must hold two contradictory beliefs. Secondly, these beliefs are held simultaneously. Thirdly, the individual is only aware of one belief: the other belief is not deliberately ignored. Finally, the nonawareness of that one belief should be motivated by the performed act.

Hogan (1983) supported the view that the individual is not aware of one belief. An underlying self-image that is held unconsciously will affect the responses in self-report questionnaires. Thus, there is no conscious semblance involved in socially desirable responding (Paulhus, 1984).

Research shows that self-deception is more severe for individuals that avoid negative thoughts, overlook minor criticisms and discount failures (Sackeim & Gur, 1979). Also, individuals that are highly conscientious will engage more in self-deceptive behaviour (Martocchio & Judge, 1997).

The concept of self-deception evokes both positive and negative reactions among psychologists (Tirole, 2002). Advocates of self-deception state that having 'positive illusions' significantly benefits the maintenance of self-esteem and self-efficacy. Contrarily, critics emphasize that people can significantly ignore negative aspects of oneself, which can cause overconfidence with all its consequences.

Tirole (2002) further elaborated on this ambiguity. Self-deception, and thereby having a selective memory, is a useful defence mechanism if one's self-confidence would be severely affected by bad news. However, it can also have a reverse effect in the form of excessive self-doubt, i.e. defensive pessimism. In that case, someone devalues his previous achievements or convinces himself in advance that he will not succeed.

Both sides of self-deception provided by Tirole (2002) can occur in dieting behaviour. Although the impact on the self-esteem of the individual is differently, both have negative effects on the individual's dieting success. In the first case, self-deceptive behaviour indicates that someone believes that he had consumed less than he actually did (Muhlheim, Allison, Heshka, & Heymsfield, 1998). This way, one can maintain his self-esteem, but will probably not lose weight. In the case of self-doubt, an individual will think he cannot meet the rules he has imposed on himself and will therefore give up, thereby being a self-fulfilling prophecy.

2.1.2 Impression management

The second component of social desirability bias is called impression management or other-deception. Impression management refers to the tendency to present a favourable image to others (Roth, Snyder, & Pace, 1986). As opposed to self-deception, impression management is done deliberately (Paulhus, 1984).

Impression management can be exerted in two ways: someone can exaggerate the presence of positive characteristics of the self, and someone can minimize the presence of negative characteristics of the self (Roth, Snyder, & Pace, 1986). In the first case, one does not attribute a new trait to oneself but exaggerates the magnitude of an existing trait. The second form of impression management is more focussed on damage control, i.e. maintaining the status quo, and avoiding punishment (Paulhus & Vazire, 2009).

In the context of self-reported dietary intake, impression management implies that one would manipulate its food intake in order to present oneself to others as healthier than he actually is (Muhlheim, Allison, Heshka, & Heymsfield, 1998). This shows resemblance with the experimenter demand effect. If an individual is told to change his food pattern and has to follow a specific diet, that person is more likely to report his food intake according to that diet, even if this does not match his actual consumption (Hebert, Clemow, Pbert, Ockene, & Ockene, 1995).

2.2 The Bayesian Truth Serum

The theory about social desirability bias shows that people are not always completely honest about their behaviour. This could be the case if people do not accurately perceive and judge a situation, while holding the belief that their behaviour is accurate (Bernardic, 2017). This can reduce the validity of self-report questionnaires, and thereby be costly for the individual himself and for others. It can preclude the individual from changing his behaviour, while it can jeopardize the information integrity for others (Bachkirova, 2016).

Wisdom of the crowd could potentially help with the elicitation of information-sensitive judgments. The wisdom of the crowd indicates that large groups of people are smarter than one individual, no matter how experienced that one person is (Surowiecki, 2004). However, it is difficult to utilize effectively if there are only subjective judgments available. It is challenging to assess the truthfulness and truth of subjective data due to the absence of external criteria that can be used to evaluate the objectiveness of the truth. Measures that employ the wisdom of the crowd, such as the majority rule or the average aggregate judgment, lack the ability to adequately incorporate variations in the quality of opinions, which could be the case for unpopular or unusual judgments (Weiss, 2009).

A method that does take these unpopular or unusual judgments into account when eliciting subjective data, is the Bayesian Truth Serum (BTS). This method makes it possible to find the truth, even if evidence consists solely of subjective judgments and these judgments might be wrong (Prelec & Seung, 2006). The wisdom of the crowd is thereby used to form a predicted distribution. By what means this is exactly accomplished and how unusual judgments are considered will be discussed in paragraph 2.2.2.

First, Bayes' theorem will be discussed, which forms a foundation on which BTS relies, and therefore is essential to review. The scoring mechanism of the BTS will be discussed in paragraph 2.2.2, followed by the assumptions that need to hold for the BTS theorem in paragraph 2.2.3. Next, previous research that has proven the effectiveness of the BTS will be examined. Finally, evidence on the relation between social desirability bias and the BTS will be reviewed.

2.2.1 Bayes' theorem

In the eighteenth century, Bayes presented a theorem that can be used when updating beliefs about the probability that an event will occur (Bayes, Price, & Canton, 1763). This theorem, also known as Bayes' rule, mainly comprises the manipulation of conditional probabilities (Olshausen, 2004). The joint probability of two events, A and B, can be expressed as:

$$P(AB) = \frac{P(A|B)P(B)}{P(B|A)P(A)}$$

where *P(A|B)* is the likelihood of *A* occurring given that *B* is true, and *P(A)* and *P(B)* are the probabilities of observing *A*, respectively *B,* independently of each other. Similarly, *P(B|A)* is the likelihood that event *B* will occur given that event *A* is true.

The following example illustrates how people should update their beliefs when new information arises. An individual holds a hypothesis. The new information can either support or contradict the given hypothesis. Therefore, the probability of rejecting the hypothesis varies with new information that emerges.

There are several scenarios possible. First, it could be that information arises that supports the hypothesis and that the hypothesis is in fact true. However, it could also be that even though the hypothesis is true, contradicting information is found. Similarly, information could emerge that supports or contradicts the hypothesis, while the hypothesis is false. These scenarios can be summarized in figure 1:

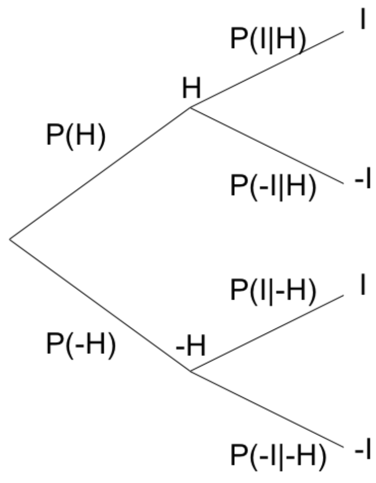


Figure 1. Bayesian updating

The letters *H* and *-H* refer to the hypothesis being true or false, respectively. The letter *I* refers to information that supports the hypothesis being true, while *-I* is information that contradicts the hypothesis being true. *P(H)* and *P(-H)* refer to the prior probabilities that the hypothesis is true or false, respectively. After the new information is presented, the posterior probability, i.e. the probability that the hypothesis is true given the information, can be expressed as:

$$P(H|I) = \frac{P(I|H) * P(H)}{P(I)}$$

where *P(I)* consists of $P(I|H) * P(H) + P(I|-H) * P(-H)$.

The probability that the hypothesis is true will change from *P(H)* to *P(H|I)* if individuals update their beliefs according to Bayes' rule.

By what means BTS theorem relies on Bayes' rule will be discussed more thoroughly in the following two paragraphs.

2.2.2 The scoring mechanism of the Bayesian Truth Serum

The BTS is a scoring method that incentivizes subjects to provide truthful answers to questions that ask for subjective information (Prelec, 2004). An answer can be regarded as truthful if it is both honest and carefully considered (Weaver & Prelec, 2013).

The scoring algorithm relies on two questions. First, respondents have to provide a personal answer about their behaviour or attitude. Secondly, they have to estimate how many other respondents will endorse the same answer. Thus, they have to predict the empirical distribution of answers.

Subsequently, the distribution of predictions and personal answers that are drawn from the same population are compared. Answers that are more common than collectively predicted, will get high scores (Prelec, 2004). Thus, answers that are 'surprisingly common' score higher. Consequently, biases that are associated with consensus-based methods are removed, and both majority and minority answers are acknowledged.

The following example about food intake illustrates how the mechanism works at the level of a single question. First, participants will have to give a personal answer to the following question:

i) Have you ever binge eaten?

a. Yes
b. No

Next, responders have to predict the distribution of the answers of the other participants. This is asked by the following question:

ii) What percentage of the other respondents will provide the same answer as you to the previous question?

The answers to the latter question are scored for how accurate they match the empirical distribution that is drawn from the population. Most interesting are the personal answers, as these answers are mainly subject to (dis)honest behaviour. These answers get an 'information score' based on whether the answer is 'surprisingly common' or not. For example, if 20 per cent of the respondents said 'Yes' to the question about whether they have binge eaten, while the predicted frequency was only 10 per cent, the answers of the first 20 per cent would be 'surprisingly common'. Hence, these answers would get a high 'information score'. Reversely, if the predicted frequency had been 40 per cent, those answers would be regarded as 'surprisingly uncommon', thereby getting a low 'information score'.

The surprisingly common criterion is based on an implication of Bayesian reasoning about population frequencies that is often overlooked. Bayesian reasoning implies that "*in most situations one should expect that others will underestimate the true frequency of one's own opinion or personal characteristic*" (Prelec, 2004). This reasoning can be deduced from the Bayesian argument that individuals that hold the same opinion will provide the highest predictions of the frequency of that opinion, as holding the opinion can be seen as valid and supportive evidence about its popularity in general (Dawes, 1990). If applied to the example question about binge eating, one should expect that a person that has ever binge eaten will estimate a higher frequency of the population that has binge eaten as well, compared to individuals who do not have the same experience. The rationale behind this behaviour is that the personal opinion forms an informative 'sample of one' (Dawes, 1989).

As mentioned in the preceding paragraph, subjects to the BTS will get an information score based on the answers they have given to both questions. The information score can be calculated as follows:

$$Information\ score = \log\left(\frac{answer's\ actual\ relative\ frequency}{(geometric)mean\ of\ answer's\ predicted\ frequency}\right)$$

A person can maximize his expected information score by answering the questions truthfully, assuming that the other respondents also behave truthfully. Therefore, telling the truth would be the most rational thing to do. Thus, with this information scoring system truth-telling is a Bayesian Nash Equilibrium (Prelec, 2004).

The BTS is similar to other Bayesian mechanisms in the sense that it makes use of the subjective correlation between one's personal beliefs and the beliefs of others (Weaver & Prelec, 2013). However, it differs from these other mechanisms when it comes to incorporating this assumption in the information score function. For example, the most popular answer will get the highest score according to the consensus scoring mechanism. Consequently, participants are incentivized to provide a dishonest answer if they believe that they hold a minority opinion. The information scoring mechanism of the BTS does not incentivize deceitful behaviour, as it does not give high scores to the most popular answer. The expected score will be lower for untruthful answers, thereby incentivizing truthful answers regardless of whether someone holds a common or uncommon belief (Weaver & Prelec, 2013).

BTS scoring is not only based on the information score that individuals get based on how truthfully they have answered the first question but also on how well they have predicted the general frequency of the population, i.e. the 'prediction score'.

The total score for responder *r* combines this prediction score with his information score and can be calculated with the following formula:

$$rScore = \sum_k \bar{x}_k^r \log\frac{\bar{x}_k}{\bar{y}_k} + \alpha \sum_k \bar{x}_k \log\frac{y_k^r}{\bar{x}_k},\ \ where\ 0 \leq \alpha$$

The first part forms the mathematical description of the total information score, whereas the second part is the total prediction score. Similar to the first formula for the information score, $\bar{x}_k$ is the actual relative frequency of answer k, while $\bar{y}_k$ is the mean of the predicted frequency of answer k. y_k^r denotes the predicted frequency of answer k by responder r. The constant α determines the weight that is given to the prediction score. If $\alpha = 1$, the formula will be symmetric and a zero-sum game (Prelec, 2004). Responders will receive one information score, as $\bar{x}_k^r$ will only take the value *1* for the answer that is endorsed by the responder. This score will be high if $\bar{x}_k$ is larger than $\bar{y}_k$, i.e. the actual frequency is higher than the average predicted frequency.

Responder r will maximize his prediction score if the individual's prediction of the frequency of answer k is equal to the actual distribution of answer k being endorsed, i.e. $y_k^r = \bar{x}_k$. Thus, the best prediction score is zero. Therefore, the prediction score can be seen as penalty for any discrepancy between the predicted frequency of individual r and the empirical distribution (Prelec, 2004). The personal answer of the responder does not influence the prediction score, while the predicted distribution by the same responder is not relevant for the information score. Thus, the answers that the responder has given to the two questions are scored independently.

2.2.3 The assumptions of the Bayesian Truth Serum

There are two assumptions that have to hold for the BTS theorem (Weaver & Prelec, 2013). First, it is assumed that individuals are Bayesian statisticians when predicting the distribution of responses to a question. People use a common prior belief as a starting point for their prediction and adjust this belief based on their own preferences, which is in accordance with Bayes' rule. Thus, individuals use all available information when updating their beliefs. An implication of this assumption is that individuals with the same preferences will predict the same posterior distribution of responses to a question.

While the BTS assumes that individuals use Bayes' rule to transform their prior belief into a posterior belief, these beliefs do not have to be elicited when using the BTS. Therefore, it is not necessary to estimate these beliefs, thereby avoiding the use of possibly non-informative priors (Bernardo, 1979). Thus, the BTS can be used in situations where the objective truth is unknown, i.e. cases where only the subjective truth can be elicited.

The second assumption on which the BTS theorem relies is that the sample has to be large enough, thereby preventing an individual prediction or response from having a significant impact on the sample results. This can be controlled by the experimenter.

2.2.4 Evidence on the effectiveness of the Bayesian Truth Serum

For truth-telling mechanisms, including the BTS, to be effective, it is essential that the subject beliefs that truth-telling will be rewarded (Prelec, 2004; Barrage & Lee, 2010). Moreover, the truth-telling incentives have to be large enough to overcome the subjects' internal incentives to behave dishonestly (Weaver & Prelec, 2013).

Several studies have tested the effectiveness of the BTS in laboratory settings.

Barrage and Lee (2010) compared the effectiveness of the BTS to two other calibration techniques, namely cheap-talk and consequentialism, in a laboratory experiment. The techniques were applied to contingent valuation surveys that are used to elicit hypothetical economic values. This type of questioning often gives rise to 'hypothetical bias', i.e. subjects often answer questions differently in a hypothetical setting compared to in a real-life situation (Cummings, Elliot, Harrison, & Murphy, 1997). Subjects were provided information about the purpose and potential efficacy of the BTS, but the theoretical foundations were omitted. In this setting, the absolute truth could not be measured. However, the researchers could observe the relative increase in honesty induced by the BTS. Even though the BTS eliminated the hypothetical bias for one of the tested goods, it was not considered the most effective truth-telling mechanism. While consequentialism was deemed most effective, the BTS had similar effects as cheap-talk. Barrage and Lee (2010) ascribed this comparability to the fact that the subjects were not familiar with the BTS and may not have understood the mechanism, thereby ignoring the idea of a monetary truth-telling incentive. Consequently, it could be that subjects deemed the idea that truth-telling would maximise their score to be cheap talk.

In the research of Weaver and Prelec (2013), subjects had to answer anonymous overclaiming questionnaires in a laboratory setting. The recognition questionnaires presented the subjects various items, such as scientific terms and brand names. Among these items several foils were included, making it possible to pinpoint whether subjects were lying and to measure the magnitude of their dishonesty. Consequently, Weaver and Prelec (2013) could test the absolute effectiveness of the BTS. They found that the BTS caused significantly more honest responses of the subjects, although it did not completely eliminate dishonesty.

Even though there are situations in which other truth-telling mechanisms may yield better results, the BTS can be regarded as an effective truth-telling mechanism in laboratory settings (Weaver & Prelec, 2013).

As dishonesty is not only relevant in laboratory settings, the effectiveness of the BTS has also been tested in online questionnaires.

John, Loewenstein and Prelec (2012) applied the BTS to an anonymous online survey regarding subjects' involvement in questionable research practices. To safeguard complete anonymity, participants could not receive their compensation themselves. Instead, their compensation was donated to a charity of their own choice, and the size of this donation depended on the truthfulness of their answers. John et al. (2012) found that the BTS had a positive effect on self-admissions of questionable research practices, which effect was even larger for methods that subjects assessed to be less defensible. Thus, subjects in the BTS group were significantly more honest than subjects that were not exposed to the BTS. A more general inference of this research could be that the BTS can be used to elicit more truthful answers to questions that provoke cognitive dissonance.

In summary, previous research shows that the BTS can be an effective tool to elicit more honest behaviour.

2.2.5 The Bayesian Truth Serum and Social Desirability Bias

Fisher (1993) proposed the indirect questioning method to mitigate the distortions caused by SDB. This technique implies that subjects have to reflect on the decisions of others. Thus, participants have to answer questions from the perspective of another person or group. The rationale behind this method is that by answering questions from the perspective of someone else, the subject will unconsciously project his own attitude towards the subject on his answer.

The research of Fisher (1993) provides evidence that indirect questioning reduces SDB in the answers to questions that are subject to social influence, but not for socially neutral questions.

The BTS can be seen as a method that combines direct and indirect questioning, as subjects have to answer both questions about themselves and about their beliefs regarding the responses of others. Thus, extrapolating the results of Fisher (1993), the BTS can be regarded an effective method to decrease the effects of SDB.

Bernardic (2017) confirmed this view by testing the impact of the BTS and Solemn Oath on self-deception in self-reporting questionnaires. Subjects that were exposed to the truth-inducing mechanisms did not only show significantly more honest behaviour, but were also less self-deceptive. These findings imply that by reducing self-deception and increasing honest answers, the BTS can improve the quality of online self-reported questionnaires.

2.3 Hypotheses

The goal of this study is to find out whether people make healthier choices after being exposed to the BTS in a dietary-related questionnaire. Based on the previously discussed literature, several hypotheses are formulated to test this relationship.

These hypotheses can be divided into two categories: eliciting honest answers and promoting healthy decisions. The first category focuses on the answers to the questions about eating behaviour, while the second category looks at the consequences for food choices. Both categories and the corresponding hypotheses will be discussed in the following paragraphs.

2.3.1 Eliciting honest answers

Based on the theory about SDB, it is expected that participants will answer the questions as if they show healthier eating behaviour than they exhibit in real life. However, the BTS can reduce the number of socially desirable answers. Consequently, it is expected that subjects are more honest about their eating behaviour if they are exposed to the BTS. This can be summarized in the following hypothesis:

H_1: Participants that are exposed to the Bayesian Truth Serum will be more honest about their eating behaviour.

2.3.2 Promoting healthy decisions

The second hypothesis focusses on the effect of the BTS on the meal options subjects choose. Participants will have to make several decisions regarding their food intake, where they have the option to choose between two alternatives: a healthy and unhealthy one. The effect of the BTS on food choices can be ambiguous.

On the one hand, people might be more aware of their, possibly unhealthy, behaviour after being treated with the BTS. The effectiveness of confronting people with the truth about their eating behaviour can be linked to the cognitive dissonance theory.

Cognitive dissonance implies that psychological discomfort arises when a person realizes that he holds two beliefs or makes two choices that are inconsistent with each other (Festinger, 1957). This discomfort can motivate people to restore consistency (Stice, Mazotti, Weibel, & Agras, 2000). Self-deception entails holding two inconsistent beliefs at the same time, while the person is only aware of one of the beliefs. Treating someone with the BTS can make the subject aware of the other belief, which leads to cognitive dissonance. In that case, behaviour shapes the person's attitude (Bertrand & Mullainathan, 2001). Thus, truth-incentivizing methods shape the person's attitude towards dietary decisions.

Other research showed that subjects that were aware of discrepancies between their behaviour and self-concept expressed guilt and self-dissatisfaction, which caused them to show more self-consistent behaviour in consecutive situations (Czopp, Monteith, & Mark, 2006).

Applied to the dietary context, if the subject considers himself to have a healthy diet but realizes that this is less the case after being treated with the BTS, he will be more likely to choose the healthy alternative in order to restore his beliefs about himself. Therefore, the BTS could incentivize people to make healthier decisions.

On the other hand, previous research shows that the BTS reduces SDB. As choosing the healthier incentive could also be a consequence of socially desirable responding, it could be that subjects in the BTS treatment group will choose the healthier reward less often.

It will be interesting to see which effect of the Bayesian Truth Serum will dominate. However, as the BTS does not eliminate SDB completely, it is expected that the effect of cognitive dissonance will prevail. Thus, the second hypothesis is as follows:

H_2: Participants that are exposed to the Bayesian Truth Serum will choose the healthy meal alternative more often than participants that are not incentivized to report honestly.

3. DATA

An empirical approach is used in this research to test whether people make healthier choices after being exposed to the BTS. The results of this research could be useful for health policy makers to tackle the problems of obesity and eating disorders. By answering questions about the behaviour of others, people could become more aware of their own behaviour and adjust their behaviour accordingly. Thus, if proven effective, BTS could work as a nudge towards healthier eating behaviour.

Ideally, a framed field experiment would have been conducted where subjects have to report their dietary intakes over a longer period of time, and some have to answer the BTS questions as well. Moreover, it would have been interesting to monitor within-subject effects. These individuals would have kept general food diaries for a couple of weeks before being treated with the BTS. This approach could detect the impact of the BTS on specific individuals. However, due to time constraints, it was not possible to conduct this ideal experiment.

As it was not feasible to conduct aforementioned framed field experiment, data is collected through a questionnaire. The aim of this survey is to collect data about people's dietary intakes and preferences, and the effect of the BTS on these reported intakes and preferences. In the next paragraph, an extensive description of the data gathering process will be given. Subsequently, the dataset will be described.

3.1 Experimental design

3.1.1 The questionnaire

A two-condition, between-subjects online questionnaire is used to collect the data. The survey is mainly distributed through social media and built with Qualtrics. The survey is subdivided into three components of questions.

First, subjects have to answer ten questions about their current food and drink intakes. These questions are based on the 'Wheel of Five' (*Schijf van Vijf*) of the Netherlands Nutrition Centre (*Voedingscentrum*) and the online test of the Dutch Dairy Association (Netherlands Nutrition Centre, 2017; Dutch Dairy Association, 2018). The questions can be subdivided into two categories: healthy and unhealthy behaviour. Examples of these two types of questions are as follows:

i) Do you eat at least two pieces of fruit per day?
   a. Yes
   b. No

ii) Do you consume soft drinks (e.g. coca cola, fanta, energy drinks) every day?
   a. Yes
   b. No

Giving an affirmative answer to the first question will indicate healthy behaviour, while affirming the latter question will illustrate unhealthy behaviour. If an individual shows dishonest behaviour, it is expected that he will confirm the healthy behaviour and deny the unhealthy behaviour as that will enhance a healthy image.

Secondly, respondents will have to make hypothetical choices about their daily food intakes. These choices are divided into five meals: breakfast, lunch, afternoon snack, dinner, and dessert. Each question consists of a healthy and unhealthy alternative. Pictures of the alternatives are provided in the survey to make sure that the participants have the same meals in mind. The motivation for these alternatives and how they are categorized as either healthy or unhealthy will be discussed in paragraph 3.1.3.

Finally, subjects are asked to provide some demographic information, such as age, gender, nationality, and highest level of education. Furthermore, participants are asked whether they are vegetarian or vegan, as this could bias the hypothetical choices of these respondents.

The complete questionnaires for both the BTS treatment group and the control group can be found in Appendix 9.5.

Respondents are randomly divided into a treatment and control group. Each respondent will be allocated to either the treatment or the control group with a probability of 50 per cent. Participants that are allocated to the control group have to answer the questions about their food intakes (such as described above) and the hypothetical choices directly, i.e. only the questions about their own preferences.

The BTS method is applied to the questionnaire of the treatment group. In addition to answering the questions about their personal dietary intakes and preferences, these subjects have to predict the distributions of the participants that will give similar answers. The following example illustrates how the question about fruit will look like for the treatment group:

i) Do you eat at least two pieces of fruit per day?

c. Yes
d. No

ii) Think about the other respondents answering this survey. What percentage of them do you feel will provide the same answer as you in the previous question?

As the questionnaire tries to elicit subjective behaviour about which someone could lie, it is important to safeguard the anonymity of the participants as much as possible. Therefore, participants have the possibility to remain completely anonymous. However, in order to make it possible to reward respondents for answering the questionnaire and behaving truthfully, they have the possibility to give up some of their anonymity by providing their e-mail adress to be possibly contacted about the prize.

The instructions on how the questionnaire and the BTS method that are provided to the respondents work, the incentives, and the hypothetical dietary choices of the questionnaire are discussed more thoroughly in the following three subparagraphs.

3.1.2 The instructions

All respondents are given general instructions about the type of questions they have to answer, the prize they can win, and the possibility of anonymity. Furthermore, responders that are in the BTS treatment group are given instructions about the BTS method. These instructions are nearly identical to the instructions given by Weaver and Prelec (2013) and are as follows:

*For each complete answer, you will earn lottery tickets based on that answer's '**Truth Score**'. Truth scoring, recently invented by a MIT professor and published in the academic journal Science* (Prelec, 2004)*, rewards you for answering truthfully. Your truth score is based on both your personal answers to each question, as well as your predictions about the answers of others. Even though only you know if you really answered truthfully or not, **people who tell the truth score higher overall**.*

You are most likely to maximize the number of lottery tickets you get if you answer every item truthfully. By 'truthfully' I mean: consider each item carefully, answer honestly, and take care to avoid mistakes.

Just by participating you already have the chance of winning the prize! For every truthful answer to the question, you will increase the chance to win. At the end, one lottery ticket will be randomly selected to be paid out for real.

According to Weaver and Prelec (2013), giving a more detailed explanation of the BTS method may do more harm than good. They found the above instructions to be sufficiently credible for respondents to believe that truth-telling will be rewarded on average. Therefore, these instructions are used in this experiment.

3.1.2 Incentives

Each respondent has the chance to win 20 euro. Whereas one of the subjects in the control group is randomly selected to win this prize, the subject with the highest BTS score will win the prize in the treatment group.

As mentioned in the theoretical framework, the BTS score is the sum of the information and prediction score:

$$rScore = \sum_k \bar{x}_k^r \log\frac{\bar{x}_k}{\bar{y}_k} + \alpha \sum_k \bar{x}_k \log\frac{y_k^r}{\bar{x}_k}, \ \ where\ 0 \leq \alpha$$

The constant α gives a particular weight to the prediction score. Similar to the study of Weaver and Prelec (2013), it is assumed that $\alpha = 0$. Therefore, the BTS score is only based on the information score. Thus, participants are only rewarded for telling the truth and not for correctly estimating the general distribution of the population.

This is emphasized in the instructions by telling the respondents that telling the truth will lead to a higher expected BTS score. The total BTS score of a respondent is the sum of the individual's information scores for each question.

To put it briefly, respondents in the control group are incentivized based on their participation, whereas subjects in the BTS treatment group are incentivized based on their performance.

3.1.3 Hypothetical choices about food

After answering questions about their current eating behaviour, participants have to make five hypothetical choices about what they prefer to eat as breakfast, lunch, afternoon snack, dinner and dessert.

A disadvantage of using hypothetical choices is that respondents possibly make different choices than they would have made in real-life situations. However, both the control and treatment group would be subject to the hypothetical bias, thereby affecting all answers of the subjects. Thus, this bias should not cause any differences between the responses of the treatment and control group.

For each meal, respondents can choose between a healthy meal and an unhealthy (or less healthy) alternative. It has been tried to limit the differences between the alternatives in the sense of taste (e.g. only sweet breakfasts, only salads for lunch, two types of 'steak' with fries for dinner), such that the personal preferences of the participants do not affect their choices excessively.

The definition of healthy and unhealthy is based on previous research and will be discussed for each meal separately.

<u>*Breakfast*</u>

For breakfast, participants can choose between (soy) yogurt with oatmeal and fresh fruit, and pancakes with maple syrup and fresh fruit.

According to the Netherlands Nutrition Centre (2018), (soy) yogurt with oatmeal and fresh fruit can be regarded as a healthy breakfast. Yogurt is high in proteins and fruit reduces the risk of cardiovascular diseases and type 2 diabetes (Netherlands Nutrition Centre, 2018). Moreover, oatmeal provides people with dietary fibres, minerals, B vitamins, carbohydrates, and protein. Oatmeal also reduces the LDL cholesterol level in the blood, thereby lowering the risk of cardiovascular diseases as well (Netherlands Nutrition Centre, 2018).

In contrast to the first breakfast option, pancakes can be regarded as unhealthy. In general, pancakes are high in saturated fat and low in dietary fibres. Refined flour is used for making pancakes, which is associated with obesity and type 2 diabetes (Spreadbury, 2012). Furthermore, sugar consumption is associated with the rise in non-communicable diseases and can have similar effects on the human body as the consumption of alcohol (Lustig, Schmidt, & Brindis, 2012).

Lunch

Respondents are presented the following two options for lunch: quinoa salad with roasted vegetables, and Caesar salad with grilled chicken.

Quinoa yields high nutritional value as it is rich in fibre, vitamins, minerals, proteins, and essential amino acids. Research shows that quinoa has a positive effect on humans' well-being, including the cardiovascular, gastrointestinal, and metabolic health (Navruz-Varli & Sanlier, 2016).

In contrast to its name, a Caesar salad cannot be regarded as a healthy lunch. The dressing is made with (raw) egg yolks, that are associated with a higher level of cholesterol and salmonella (Spence, Jenkins, & Davignon, 2010). By adding parmesan cheese and chicken to the salad, it becomes high in saturated fat. Moreover, faecal contamination in chicken meat imposes a risk for people's health (Greger, 2016).

Afternoon snack

As an afternoon snack, subjects can choose between pistachio nuts and chips.

Pistachio nuts yield several health benefits, such as maintaining healthy anti-inflammatory activity, glycaemic control, and benefiting the heart. This can be ascribed to its high density of vitamins, dietary fibre, protein, magnesium, potassium, and unsaturated fats. If moderately consumed, it can help people lose weight because of its satiation effects (Dreher, 2012).

Chips are high in saturated fats and salt. Negative effects of the consumption of dietary sodium (known as salt) are lower cardiovascular health and raised blood pressure (Brown, Tzoulaki, Candeias, & Elliot, 2009). Furthermore, toxic acrylamide is found in fried potatoes, which increases the risk of cancer (Netherlands Nutrition Centre, 2018).

Dinner

Respondents can choose between two types of 'steak frites' for dinner: cauliflower steak with sweet potato fries, and steak with French fries.

Similar to most vegetables, cauliflower is high in dietary fibre and vitamins, and reduces the risk of cardiovascular diseases (Netherlands Nutrition Centre, 2018). Sweet potato has benefits for the health as it is high in vitamins and minerals, and scores lower on the glycaemic scale than normal potatoes, thereby being better in maintaining the blood sugar level (Bovell-Benjamin, 2007).

According to the World Health Organisation, eating red meat is linked to several types of cancer, including colorectal, pancreatic, and prostate cancer (World Health Organization, 2015; Gallagher, 2015). Even though potatoes yield health benefits, frying them abolishes these benefits. French fries are high in fat and acrylamide, thereby contributing to the risk of obesity and getting cancer (Camire, Kubow, & Donnelly, 2009).

Dessert

Finally, participants have to choose between fresh strawberries and strawberry cheese cake as dessert.

Eating at least 200 grams of fruit per day lowers the risks of type 2 diabetes, several types of cancer, such as colon and lung cancer, and cardiovascular diseases (Netherlands Nutrition Centre, 2018). Strawberries are high in antioxidants, vitamin C, and folic acid (Van Duyn & Pivonka, 2000).

Strawberry cheesecake is high in sugar and saturated fat, thereby making it an unhealthy dessert. As most unhealthy meals, it can result in heart diseases and type 2 diabetes (Doll, et al., 2009).

3.2 Description of the dataset

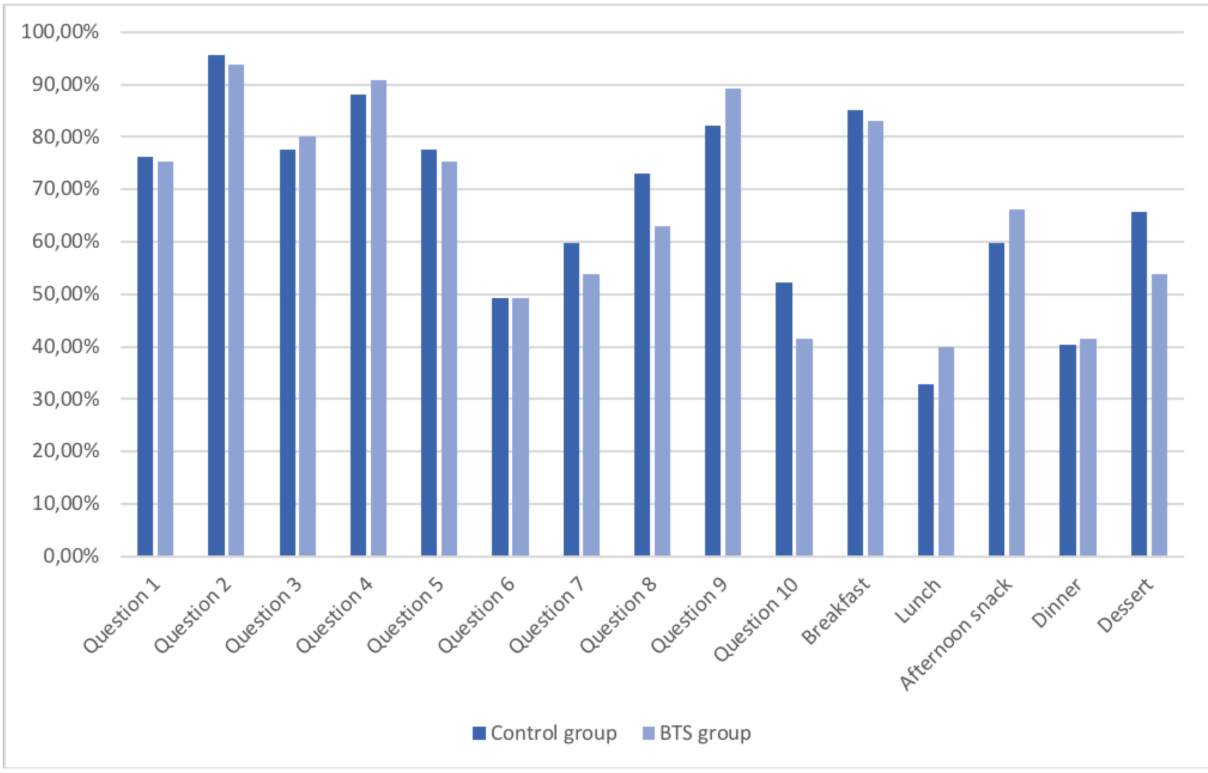
The dataset contains 132 independent observations at the subject level. Detailed descriptive statistics can be found in appendix 9.2.1. On average, subjects are 36.30 years old, while the youngest participant is 13 years old, and the oldest participant is 78 years old. 93 females answered the questionnaire, while 39 males filled out the survey. The majority of the sample is from the Netherlands (95.45%). The other participants are from Belgium, Bulgaria, China, Lebanon and the UK. 36.36% of the participants finished a Master degree at university, while 28.79% obtained a bachelor degree, and 2.27% a doctorates degree. The highest education level of 18 participants (13.64%) is college, 23 participants (17.42%) finished high school and 2 participants (1.52%) finished primary school so far. 6.06% of the responders follows a vegetarian or vegan diet.

In appendix 9.2.2 the descriptive statistics per treatment group can be found. This information is relevant for checking the homogeneity of the two groups. 65 responders were exposed to the BTS treatment, whereas 67 responders filled out the control questionnaire.

The average age of the subjects in the control group is higher by one year than the average age of the BTS treatment group. Females form the majority in both groups, but are more prevalent in the BTS treatment group (73.85%) than in the control group (67.16%). The distribution of Dutch participants is fairly similar across both groups (95.52% in the control group versus 94.38% in the BTS group), while the other nationalities are only represented in either one of the two groups. In both groups, participants with a bachelor and master degree as highest obtained education level form the majority (65.67% of the control group and 64.61% of the BTS group). Finally, vegans or vegetarians represent approximately 6% of the participants in both groups.

Considering this demographic information, the control and treatment group seem homogeneous.

Appendix 9.2.3 displays the frequencies and percentages of the healthy answer to each question. This table is summarized in graph 1, which can be found on the next page. Which answers are regarded as healthy is described in appendix 9.1 (description of the dataset 2).



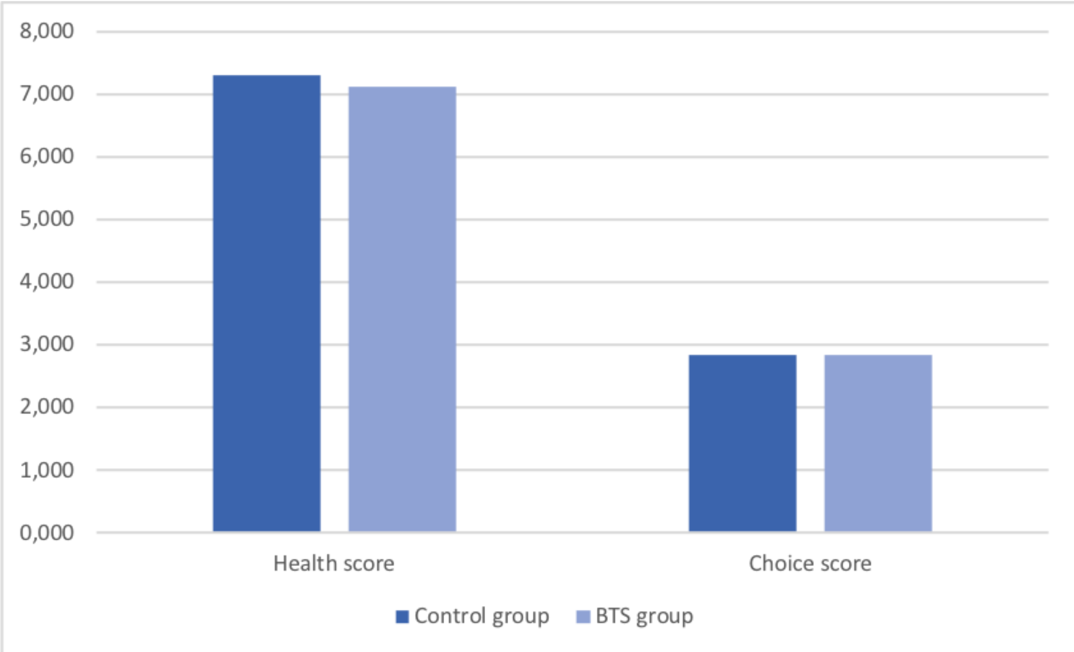
Graph 1. Frequencies of healthy answers per treatment group

Even though the differences between most percentages is low, subjects in the control group give the healthy answer more often to six questions about their current food and drink intakes. The percentages are nearly equal for question 6, which asks whether the individual eats at least 2 pieces of fruit per day. Subjects in the BTS treatment group show healthier behaviour for questions 3, 4, and 9 than the subjects in the control group.

Regarding the meal choices, subjects in the BTS treatment group choose the healthy option more often for lunch, afternoon snack, and dinner, while they prefer the unhealthy option more often for breakfast and dessert. The differences in percentages are more pronounced for the choice part of the questionnaire than for the current intakes part.

Based on the answers to the questionnaire, two scores can be determined: The ‘health score’, which displays how healthy the diet of the participant is based on his current drink and food intakes, and the ‘choice score’, which illustrates the number of healthy choices the participant makes. The health score can take the values 0 to 10, as participants get 1 point for each healthy answer that they give. The choice score can take the values 0 to 5, depending on how often the participant chooses the healthy meal option.

On average, subjects in the control group have a health score of 7.313 points and a choice score of 2.836 points. For the BTS treatment group, the average health score is 7.123 points and the average choice score 2.846 points. Graph 2 illustrates the mean scores per treatment group.



Graph 2. Mean Health score and Choice score per treatment group

This suggest that even though subjects in the BTS treatment group have a lower health score, they make healthier hypothetical choices. However, the standard deviations are higher for the BTS treatment group and the differences are small. Whether these differences are significant and how this will be tested, will be discussed in the results and methodology respectively.

In the next chapter, the methodology that is used to analyse these data will be discussed.

4. METHODOLOGY

To test the hypotheses several analyses will be performed. First, the methodology that is used to assess whether the common prior assumption holds in this research will be explained. In paragraphs 4.2 and 4.3, the tests that are used to analyse the data will be described for hypothesis one and hypothesis two respectively.

Non-parametric tests are used to analyse the data, because the underlying population distribution is unknown and the data consists of independent observations. The advantages of using this type of tests are that the variables do not have to be measured in an interval scale, a small sample size can be used, and outliers have less impact on the results. However, these tests are less powerful than parametric tests, making it harder to reject the null-hypothesis when it is false. Therefore, parametric tests are used as robustness check. The OLS and Probit regression models test whether the nonparametric results hold for parametric tests as well.

4.1 The common prior assumption of the BTS

As discussed in the theoretical framework, it is assumed that people use their common prior belief as a reference point for predicting the distribution. This implies that individuals that confirm healthy behaviour in their current food and drink intake will give a higher prediction rate of healthy behaviour than the individuals that do not give the healthy answer.

It is tested whether this assumption holds for questions one to ten about the respondents' current food and drink intakes and the meal choices by comparing the mean prediction rates. The mean estimated percentages of respondents that have answered 'Yes' to the question according to the individuals that have answered 'Yes' and the individuals that answered 'No' to that question are compared. If those means are consistently higher amongst the individuals that answered 'Yes', it confirms that the common prior assumption holds. The same method is used to compare the mean prediction rates of the healthy meal choices.

4.2 Hypothesis one: the BTS and (dis)honesty about current dietary intakes

The first hypothesis tries to answer the question whether participants that are exposed to the BTS will be more honest about their current food and drink intakes. In order to test this, a Fisher's exact test will be used for each question concerning their current diet. The null hypothesis of this test is that both possible answers (i.e. 'Yes' and 'No') to the question are evenly distributed over the control and treatment group. If the p-value is lower than the significance level, the null hypothesis can be rejected. If this is the case, there is evidence that the allocation of a subject to the BTS treatment affects his answer to the question.

Furthermore, a Mann-Whitney U test is run to determine whether there is a significant difference in the health score of the control and treatment group. The null-hypothesis is that the mean health scores of both groups are equal. If this hypothesis can be rejected, it shows that the average health score of the BTS treatment group is significantly different from the health score of the control group. As appendix 9.2.4 already shows that the BTS treatment group has a lower average health score than the control group, rejecting the null-hypothesis would imply that the BTS treatment group scores lower, i.e. admits more unhealthy behaviour.

Even though it is unknown whether the data satisfies the assumptions of parametric tests, three regression models are used to examine the relation between the treatment and the answers to the questions about the respondents' current dietary intakes.

First, a binomial Probit regression model is performed for each question, using the binary variables of the respective question as the dependent variable, and the treatment as independent variable. Age and gender are added as control variables. The model does not control for being vegan or vegetarian, nationality and education level, as these variables predict the probabilities of some answers perfectly. Therefore, these observations should be dropped. To estimate the probability that the BTS affects the answers of all participants, it is chosen to exclude these variables. The dependent variable, i.e. the answer to the respective question, takes the value 1 if the healthy answer is given and 0 otherwise. By running this regression model, it can be examined whether being in the BTS treatment group affects the probability that the respondents will give the healthy answer.

Secondly, an OLS regression model is used to estimate the influence of the treatment on the magnitude of the health score. The model controls for age, gender, nationality, education level and whether the individual is vegan/vegetarian or not. As nationality and education level both consist of six categorical dummy variables, for each category one dummy variable is omitted to avoid multicollinearity. It is chosen to omit the nationality 'UK' and the education level 'primary school'.

As an extension of the OLS regression model, an ordered Probit regression model is employed to estimate the ceteris paribus effect of the BTS and the control variables on the probability of having a certain health score. As no participant has a health score of 1 point or lower, the ceteris paribus effects are only calculated for health scores of 2 to 10 points. This gives a more detailed view of the effects on the health score.

4.3 Hypothesis two: the BTS and making healthier choices

The second hypothesis concerns the effectiveness of the BTS on the hypothetical food choices participants make. Similar to the testing of hypothesis one, Fisher's exact test will be employed for each hypothetical choice to analyse whether the distribution of these choices is significantly different for the BTS treatment group.

Next, a Mann Whitney U test is run to assess whether the combined choice score is significantly different for the BTS treatment group. If the null-hypothesis, i.e. the mean choice scores of both groups are equal, can be rejected, it provides evidence that one of the groups makes significantly healthier choices.

Additionally, three types of regressions are performed to examine the relation between the hypothetical choices and the treatment, and to check for the robustness of the nonparametric results.

First, a binomial Probit regression model is performed for each hypothetical choice, using the binary variables of the respective choice as the dependent variable. Treatment, health score, the interaction effect of the treatment and health score, age and gender are used as independent variables. By adding the first two independent variables, it can be examined whether BTS and the personal health score affect the hypothetical choices. The interaction effect captures the difference in health scores of the BTS treatment group and the control group and can therefore show whether there is a significant difference between the two groups.

Secondly, an OLS regression model is used to estimate the influence of the treatment on the choice score. The explaining variables are BTS, health score, the interaction effect between treatment and health score, vegan, age, gender, nationality and education level. The variable vegan controls for the fact that vegetarians or vegans had less choices to make as some choice options contained animal products. Similar to the OLS regression model for hypothesis one, the dummy variables ‘UK’ and 'primary school’ are omitted to avoid multicollinearity for the categorical variables nationality and education level respectively.

Finally, an ordered Probit model is run to assess the directional effect of the BTS on the probability of having a certain choice score. This gives a detailed view on the coefficients of the OLS regression model. The choice score is used as the dependent variable, which can take the values 0 to 5. Treatment, health score, the interaction effect between treatment and health score, and the demographic variables are used as independent variables.

5. RESULTS

First, the results that indicate whether the common prior assumption of BTS holds are discussed. Subsequently, the results are reviewed per hypothesis. For hypothesis one, Fisher's exact tests, a Mann-Whitney U test, and OLS, Binominal, and Ordered Probit regression models are used. Hypothesis two is analysed by the same tests and regression models but with different variables.

A 5% significance level will be maintained to interpret the nonparametric results. The complete values can be found in the Appendix.

5.1 Testing the common prior assumption of the BTS

As previously mentioned, the BTS assumes that people that give a certain answer will give a higher prediction of the distribution of that answer relative to those that give the alternative answer. The results of testing whether this assumption holds can be found in table 1. The two possible answers are named 'A' and 'B'. 'A' corresponds with 'Yes' for the ten questions about the current dietary intakes and with the healthy alternative for the five meal choices, whereas 'B' corresponds with 'No' and the unhealthy choice. In column 3, the average prediction of answer A by individuals that responded A and those who responded B are given in percentages. The difference between those two predictions is provided in column 4.

The testing of the assumption can be illustrated by the answers to question one. The individuals that responded 'Yes' predicted that 70.08% would answer 'Yes', while those who answered 'No' predicted that 56.38% would answer 'Yes'. Thus, the difference in their predictions of the prevalence of 'Yes' is 13.71%. A positive difference in the predictions of the two groups indicates that the common prior assumption of the BTS holds. As shown in table 1, this is the case for all questions except question two, which refers to using sugar in your coffee or tea.

Question	Answer	Average prediction of A %	Average prediction of A% from 'A' respondents minus average prediction of A% from 'B' respondents
1	A – Yes	70.08	13.71
	B – No	56.38	
2	A – Yes	32.25	-17.19
	B – No	49.44	
3	A – Yes	52.56	16.33
	B – No	36.23	
4	A – Yes	53.17	7.06
	B – No	46.10	
5	A – Yes	63.06	10.29
	B – No	52.78	
6	A – Yes	47.50	11.53
	B – No	35.97	
7	A – Yes	77.40	26.43
	B – No	50.97	
8	A – Yes	52.37	12.45
	B – No	39.92	
9	A – Yes	56.00	5.66
	B – No	50.34	
10	A – Yes	74.84	11.47
	B – No	63.37	
Breakfast	A – Healthy	59.30	14.21
	B – Unhealthy	45.09	
Lunch	A – Healthy	43.46	6.72
	B – Unhealthy	36.74	
Snack	A – Healthy	45.23	6.82
	B – Unhealthy	38.41	
Dinner	A – Healthy	38.15	10.67
	B – Unhealthy	27.47	
Dessert	A – Healthy	51.54	18.04
	B – Unhealthy	33.50	

Table 1. The results of testing the common prior assumption of BTS

5.2 The relation between BTS treatment and being honest about current dietary intakes

5.2.1 Nonparametric tests: Fisher's exact tests and Mann Whitney U test

As discussed in the methodology, Fisher's exact tests are employed for each question about the participant's current dietary intakes to test whether the BTS induces more honest answers.

The theory predicts that participants would give socially desirable answers to the questions, thereby showing healthier eating behaviour than they truthfully exhibit. Previous research shows that the BTS reduces the extent to which subjects give socially desirable answers. This implies that respondents in the BTS treatment group would give different answers than participants in the control group.

This expectation corresponds with the null-hypothesis that the proportion of the answers 'Yes' and 'No' is the same for the BTS treatment group and the control group. For each question, the null-hypothesis cannot be rejected as all p-values are above the significance level (Appendix 9.3.1, results 1-10). The corresponding p-values of each Fisher's exact test are summarized in table 2.

Question	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
p-value	1.00	0.72	0.83	0.78	0.84	1.00	0.60	0.26	0.32	0.23

Table 2. The p-values of the Fisher's exact tests for questions one to ten

Thus, there is no evidence that the BTS treatment affects participants' answers to the questions about his current food and drink intakes.

Based on the answers to questions one to ten, respondents have received a health score. A Mann-Whitney U test is used to indicate whether the health scores of the control group and the BTS treatment group differ. The null-hypothesis that the mean health scores of both groups are the same cannot be rejected ($p = 0.79$). Thus, the average health scores of both groups do not differ significantly (Appendix 9.3.1, result 11). As the descriptive statistics of appendix 9.2.4 indicate that the health score of the BTS treatment group is lower than that of the control group, this result implies that the BTS treatment group does not score significantly lower, i.e. does not significantly admit more unhealthy behaviour.

These results are not in line with the theory as it was expected that the BTS treatment would lead to different answers to the ten questions and correspondingly, a lower health score.

5.2.2 Parametric tests: OLS and Probit regression models

For each question, a binomial Probit regression model is performed, although it is not assured that the assumptions of the parametric test hold. The results of the individual regressions, including the complete parameter values, standard errors and p-values, can be found in Appendix 9.4.1 (results 18-37). These results are merged and summarized in table 3.

	Q1 (Yes=1)	**Q2 (No=1)**	**Q3 (Yes=1)**	**Q4 (No=1)**	**Q5 (No=1)**	**Q6 (Yes=1)**	**Q7 (No=1)**	**Q8 (Yes=1)**	**Q9 (No=1)**	**Q10 (No=1)**
BTS	-0.090	-0.167	0.098	0.093	-0.119	-0.030	-0.248	-0.282	0.341	-0.256
Age	0.026***	0.003	0.004	-0.013	0.025***	0.015**	0.020***	0.001	0.022**	-0.008
Male	-0.300	0.033	0.337	-1.135 ***	-0.451*	-0.212	-0.807 ***	0.026	-1.025 ***	0.181
Constant	-0.025	1.597***	0.516	2.179***	0.091	-0.488	-0.168	0.583*	0.641	0.279
Pseudo R^2	0.083	0.005	0.012	0.153	0.097	0.034	0.116	0.009	0.187	0.020
Log likelihood	-67.041	-27.247	-67.412	-37.835	-65.003	-88.392	-79.839	-81.785	-44.202	-89.402
Obs.	132	132	132	132	132	132	132	132	132	132

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Table 3. Summary of the Probit regression models for questions one to ten

The average marginal effects of each parameter are calculated to interpret the magnitude of the parameters. The complete average marginal effects analyses can be found in Appendix 9.4.1 (results 18-37). Again, the individual results are merged and summarized in table 4.

	Q1 (Yes=1)	Q2 (No=1)	Q3 (Yes=1)	Q4 (No=1)	Q5 (No=1)	Q6 (Yes=1)	Q7 (No=1)	Q8 (Yes=1)	Q9 (No=1)	Q10 (No=1)
BTS	-0.025	-0.018	0.028	0.015	-0.033	-0.011	-0.085	-0.100	0.062	-0.100
Age	0.007***	0.000	0.001	-0.002	0.007***	0.006**	0.007***	0.000	0.004**	-0.003
Male	-0.085	0.004	0.097	-0.178 ***	-0.125*	-0.082	-0.278 ***	0.009	-0.187 ***	0.070
Obs.	132	132	132	132	132	132	132	132	132	132

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Table 4. Summary of the average marginal effects analyses for the answers to question one to ten

According to table 4, age has a significant positive effect on the probability of giving the healthy answer to question 1 (0.7 percentage points), question 5 (0.7 percentage points), question 6 (0.6 percentage points), question 7 (0.7 percentage points) and question 9 (0.4 percentage points), ceteris paribus.

Moreover, being a male significantly decreases the probability of giving the healthy answer to question 4 by 17.8 percentage points, to question 7 by 27.8 percentage points, and to question 9 by 18.7 percentage points at a 5% significance level, ceteris paribus. The probability of giving the healthy answer to question 5 decreases by 12.5 percentage points if the individual is male, ceteris paribus, given a significance level of 10%.

As the effect of BTS on the probability of giving a healthy answer is insignificant for all questions, and differs in sign, no conclusions can be made about this effect. Thus, no concluding remarks can be made about the effect of BTS treatment on the probability of giving the healthy answer to questions one to ten.

Next, an OLS regression model is run to estimate the effect of the BTS and control variables on the health score, which combines the answers to questions one to ten. This model can be found in appendix 9.4.1 (result 38). As the effect of BTS treatment on the health score is insignificant, it is impossible to deduce any statistical conclusions from it. However, the OLS model indicates that BTS decreases the health score, which has the same direction as was expected according to the theory.

Being vegan or vegetarian significantly increases the health score with 1.342 points, ceteris paribus. An increase of 1 year in age significantly increases the health score with 0.032

points, ceteris paribus. Having the Belgian or Chinese nationality has a positive significant effect on the health score by 4.363 points and 5.026 points respectively, ceteris paribus. Being Dutch increases, on average, the health score with 3.002 points, at a 10% significance level and ceteris paribus.

Subsequently, an ordered Probit regression model is used to assess the ceteris paribus effect of the BTS and the demographics on the probability of having a certain health score. The Probit model and the average marginal effects analysis can be found in appendix 9.4.1 (results 39 and 40).

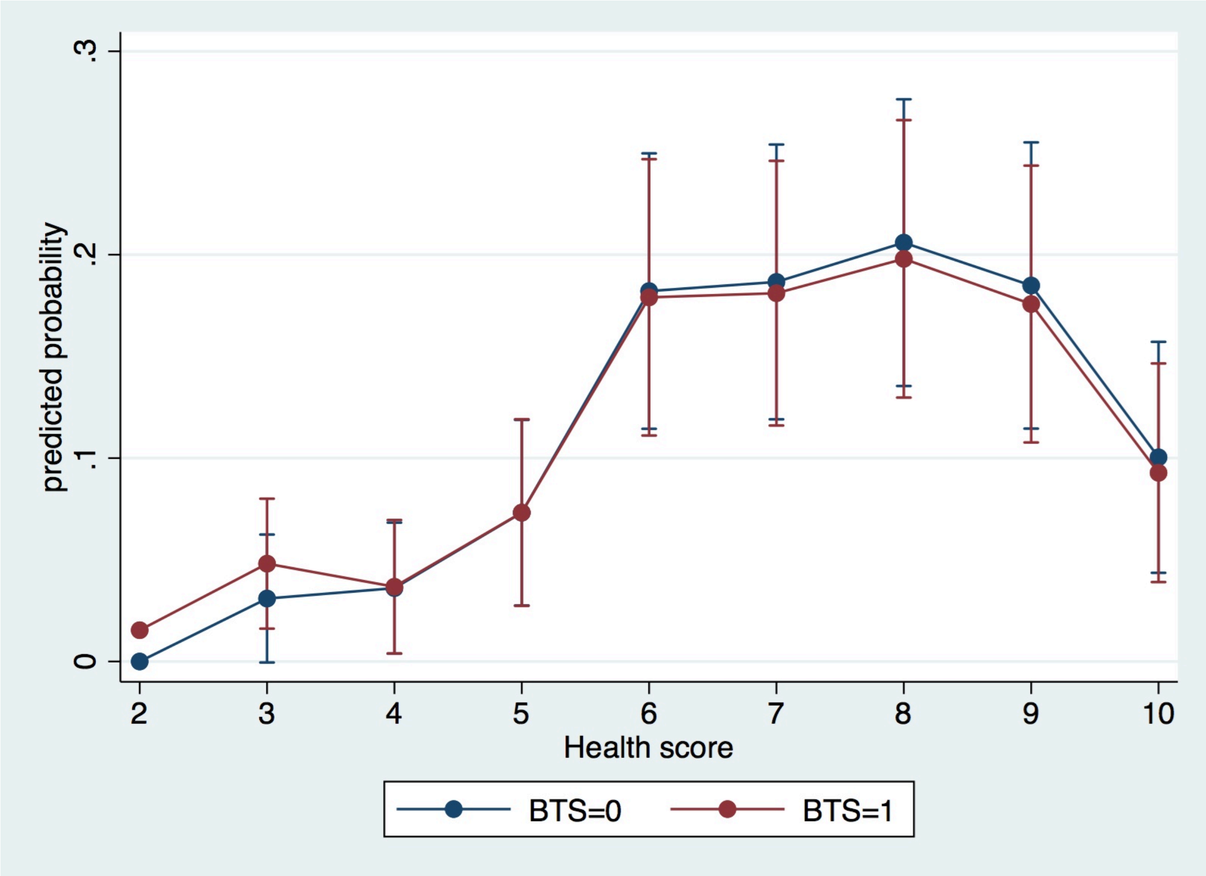
Being vegan or vegetarian decreases the probability of having a health score of 5 and 6 points with 7.1 and 10.2 percentage points respectively, ceteris paribus. The probability of having a health score of 9 and 10 points increases with 12.5 and 13.1 percentage points respectively if the participant is vegan or vegetarian, ceteris paribus.

Age has a significant effect on the probability of having a health scores of 3 to 10 points. This effect is negative and 0.1 or 0.2 percentage points on the probability of having a health score of 3 to 7 points, whereas it has positive effect of 0.1 to 0.3 percentage points on the probability of having a health score of 8 to 10 points, ceteris paribus.

If the participant is male, the probability of having a health score of 5 and 6 points increases with 3.8 and 5.4 percentage points respectively, ceteris paribus. However, the effect is significantly negative on the probability of having a health score of 8 to 10 points, ceteris paribus. This indicates that men are less likely to get a health score of 8 points or higher than women.

The average marginal effects of BTS are insignificant and can therefore not be statistically interpreted. However, the signs indicate that BTS has a positive effect on the probabilities of having a health score of 2 to 7 points, while the effect is negative for the probabilities of having a health score of 8 to 10 points.

The marginal effect of being in either the treatment or control group on the health score is illustrated in graph 3 (see Appendix 9.4.1, result 41 for the corresponding table).



Graph 3. Predictive margins of BTS on the Health score, with a confidence interval of 95%

The graph indicates that the average probability of having a health score of 2 and 3 points is higher if the participants were treated as if they were exposed to BTS, whereas the reversed effect occurs for health scores of 6 points and higher. Even though no reliable conclusions can be made about the difference between the predicted probabilities, this graph suggests that BTS treatment results in lower health scores than no treatment. This is in line with the expectation that BTS would result in less socially desirable responding and thereby, less healthy reported behaviour.

5.3 The relation between BTS treatment and hypothetical food choices

5.3.1 Nonparametric tests: Fisher's exact tests and Mann Whitney U test

According to the theory, it is expected that the respondents of the BTS treatment group will choose the healthy alternative more often than the participants of the control group. Similar nonparametric tests are used to test whether BTS affects the hypothetical food choices of the respondents.

First, five Fisher's exact tests are run to test the distribution of the alternatives of each meal choice separately. The null-hypothesis that the distribution of the healthy and unhealthy alternative is the same for both groups cannot be rejected for any meal choice, as all p-values are higher than the significance level of 5% (Appendix 9.3.2, results 12-16). The p-values of these five Fisher's exact tests are presented in table 5.

Meal	**Breakfast**	**Lunch**	**Snack**	**Dinner**	**Dessert**
p-value	0.82	0.47	0.48	1.00	0.21

Table 5. The p-values of the Fisher's exact tests for the five meal choices

Hence, these results indicate that BTS treatment does not affect the hypothetical meal choices of the participants.

Next, a Mann-Whitney U test is run to test whether BTS treatment yields a different choice score. This choice score consists of one point for each healthy choice, with a maximum of five points. The null-hypothesis that the mean choice scores of the two groups are identical cannot be rejected ($p = 0.88$). Thus, there is no evidence that the choice scores of the control and treatment group differ significantly (Appendix 9.3.2, result 17). Even though the descriptive statistics of appendix 9.2.4 illustrate that the average choice score of the BTS treatment group is slightly higher than the choice score of the control group, this difference is insignificant, which implies that the BTS treatment group did not choose the healthy alternative more often.

The results of the nonparametric tests contradict the expectations based on the theory, i.e. BTS treatment would result more often in healthier choices and subsequently, a higher choice score.

5.3.2 Parametric tests: OLS and Probit regression models

Six binomial Probit regression models are performed to see whether the nonparametric results for the hypothetical food choices also hold under parametric conditions. The complete output per meal choice, including the parameter values, standard errors, and p-values, can be found in Appendix 9.4.2 (results 42-51). Table 6 contains the merged and summarized results for all the meal choices.

	Breakfast (Healthy=1)	**Lunch (Healthy=1)**	**Snack (Healthy=1)**	**Dinner (Healthy=1)**	**Dessert (Healthy=1)**
BTS	-2.018*	-0.111	-0.905	1.053	-0.494
Health score	0.013	-0.035	0.180*	0.243**	0.048
BTS*Health score	0.296*	0.035	0.167	-0.144	0.024
Age	0.045***	-0.005	0.026***	-0.005	0.008
Male	-0.006	-0.929***	0.388	-0.845***	-0.132
Constant	-0.411	0.247	-2.088***	-1.595**	-0.187
Pseudo R^2	0.232	0.072	0.188	0.109	0.032
Log likelihood	-44.419	-80.256	-70.669	-79.534	-86.064
Obs.	132	132	132	132	132

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Table 6. Summary of the Probit regression models for the hypothetical meal choices

The magnitude of the parameters can be interpreted by using an average marginal effects analysis. The detailed results per hypothetical meal choice can be found in Appendix 9.4.2 (results 42-51), which are merged and summarized in table 7.

	Breakfast (Healthy=1)	Lunch (Healthy=1)	Snack (Healthy=1)	Dinner (Healthy=1)	Dessert (Healthy=1)
BTS	-0.376*	-0.038	-0.274	0.363	-0.184
Health score	0.002	-0.012	0.054*	0.084**	0.018
BTS*Health score	0.055*	0.012	0.050	-0.050	0.009
Age	0.008***	-0.002	0.008***	-0.002	0.003
Male	-0.001	-0.322***	0.117	-0.291***	-0.049
Obs.	132	132	132	132	132

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Table 7. Summary of the average marginal effects analyses for the hypothetical meal choices

Table 7 shows that BTS has a significant negative effect of 37.6 percentage points on the probability of choosing the healthy breakfast alternative, albeit at a 10% significance level, ceteris paribus. This contradicts the expectation based on the theory. The sign of the effect of BTS is negative for lunch, snack and dessert choice as well, while it is positive for the dinner choice. However, as these results are insignificant, no concluding remarks can be made about the effect of BTS treatment on the probability of making a healthy meal choice.

An increase in the health score significantly increases the probability of choosing the healthy snack alternative by 5.4 percentage points and the probability of choosing the healthy dinner alternative by 8.4 percentage points, given a 10% and 5% significance level respectively, ceteris paribus.

The coefficient of the intercept of BTS and the health score for the breakfast choice indicates that the probability that the individual chooses the healthy breakfast option significantly increases with 5.5 percentage points for an increase of 1 point in the health score of an individual in the BTS treatment group, ceteris paribus, albeit at a 10% significance level.

Furthermore, an increase of 1 year in age significantly increases the probability of choosing the healthy breakfast and snack alternative by 0.8 percentage points, ceteris paribus.

Male participants are 32.2 percentage points less likely to choose the healthy lunch alternative and 29.1 percentage points less likely to choose the healthy dinner option, ceteris paribus.

The signs of the coefficients of the variables age and male are consistent with the results of the Probit models regarding the questions about current dietary intakes.

Subsequently, an OLS regression model is employed to estimate the effect of BTS treatment and the control variables on the choice score. Participants were given 1 point for each healthy choice, resulting in a minimum choice score of 0 points and a maximum choice score of 5 points. The OLS model for the choice score is displayed in appendix 9.4.2 (result 52).

The model indicates that BTS had a negative effect on the choice score, which differs from the expectation based on the theory. However, as this effect is insignificant, no statistical conclusions can be deduced from it.

An increase of 1 point in health score significantly increases the choice score with 0.166 points, ceteris paribus.

Moreover, the choice score of male participants is on average 0.476 points lower than the choice score of females, ceteris paribus. The variables vegan and age have a significant positive effect of 0.770 and 0.012 points respectively, at the 10% significance level and ceteris paribus.

Finally, an ordered Probit regression model is run to show the effect of the BTS, health score, the interaction between the treatment and the health score, and the demographics on the probability of having a certain choice score. The results of this model are presented in appendix 9.4.2 (result 53). To interpret the magnitude of the coefficients, an average marginal effects analysis is performed. The average marginal effects of the independent variables on the probability of having a certain choice score can be found in appendix 9.4.2 (result 54).

An increase of 1 point in health score significantly decreases the probability of having a choice score of 1 and 2 points by 3.1 and 2.3 percentage points respectively, ceteris paribus. The probability of having a choice score of 4 and 5 points significantly increases by 3.0 and 2.2 percentage points respectively, if the health score increases by 1 point, ceteris paribus. These results seem reasonable, as it can be expected that if someone eats healthy, he will make healthier choices as well.

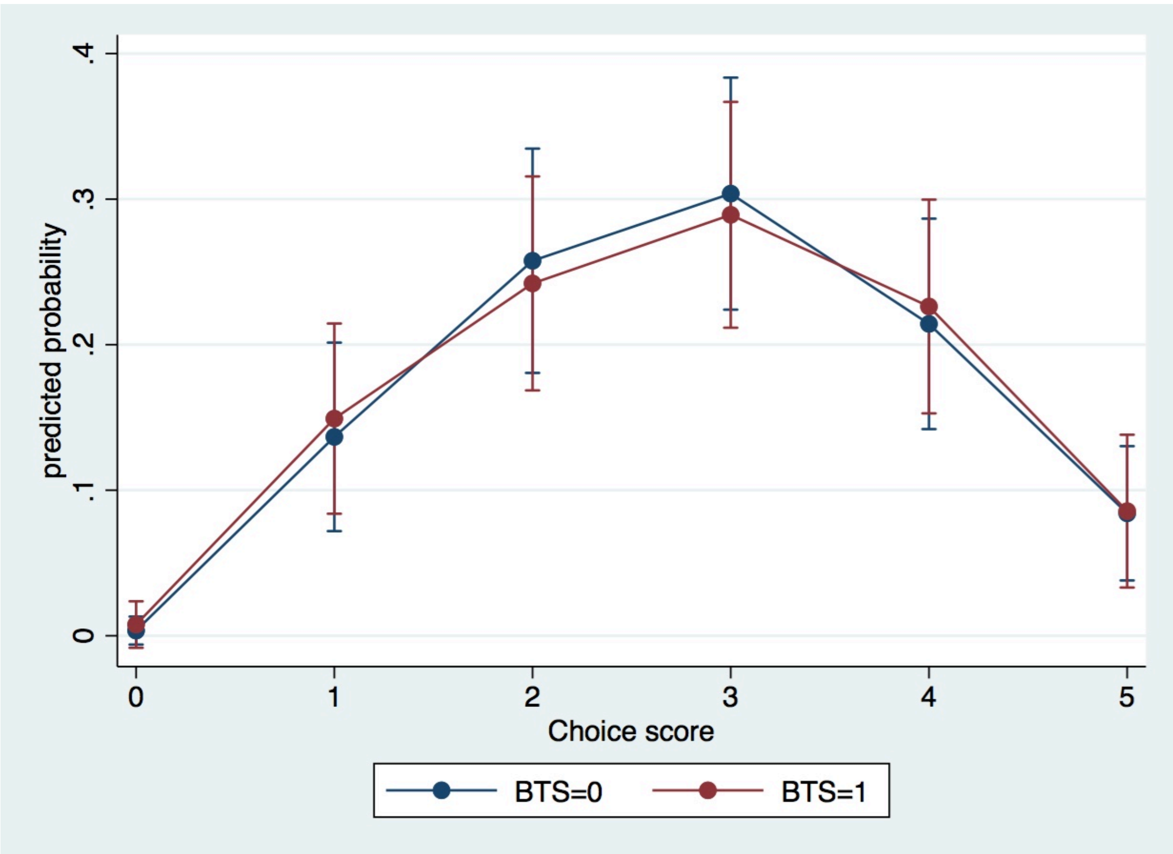
Being vegan or vegetarian significantly decreases the probability of having a choice score of 1 (14.8 percentage points) and 2 (10.9 percentage points) points, while the effect is positive on the probability of having a choice score of 4 (14.0 percentage points) and 5 (10.3 percentage points) points, albeit at a 10% significance level, ceteris paribus.

Similar to the effect on the health score, an increase in one year of age decreases the probability of having a choice score of 1 and 2 points by 0.2 percentage points at a 10% significance level, ceteris paribus. The sign of the effect switches for the probability of having a choice score of 4 (0.2 percentage points) and 5 (0.1 percentage points) points, ceteris paribus.

Being male significantly increases the probability of having a choice score of 1 and 2 points by 8.9 and 6.5 percentage points respectively, whereas it significantly decreases the probability of having a choice score of 4 and 5 points by 8.4 and 6.2 percentage points respectively, ceteris paribus.

Similar to the results of the OLS regression model, BTS has no significant effect on the choice score. Nevertheless, the average marginal effects analysis indicates that this effect is positive on the probability of having a choice score of 0 to 2 points, whereas the effect is negative on the probability of having a choice score of 3 to 5 points. This implies that the BTS did not induce healthier choices.

The effect of being in either the treatment or control group on the predicted probability of having a certain choice score is represented in graph 4 (see Appendix 9.4.2, result 55 for the corresponding table).



Graph 4. Predictive margins of BTS on the Choice score, with a confidence interval of 95%

The graph suggests that the effect of being treated with the BTS compared to no treatment can be ambiguous, as the predicted probability of having a low choice score (< 2 points) or high choice score (> 3 points) is higher if the participants were treated as if they were exposed to BTS, although no statistical conclusions can be drawn from it.

6. DISCUSSION

As discussed in the results, no evidence is found that supports the hypotheses. This could be the result of several limitations that this study faces. Based on some of these limitations, recommendations for further research can be made.

Firstly, there are several limitations to the questionnaire design. The anonymity of the respondents could not be completely safeguarded as their e-mail addresses were needed to possibly inform them about the prize. Also, there could possibly be a selection bias in the data, as it is more likely that acquaintances have filled in the questionnaire. Besides possibly biasing the results, this could also jeopardise the feeling of anonymity. The prospect of complete anonymity can be important for eliciting truthful subjective data. Therefore, there is a chance that the answers of the participants were influenced by the lower level of anonymity. A possible solution would be that the prize money is donated to a charity of choice. In that case, no e-mail address is required. However, this solution has some pitfalls as well, as it depends on altruistic behaviour of the participants.

Another pitfall of the questionnaire is that the chance of winning real money is small and the monetary amount might be too low, thereby limiting the extrinsic motivation of the respondents to report their true behaviour.

Secondly, the meal choices the respondents had to make were not perfect. People might hold different beliefs on what is healthy and what is not. This could bias the decisions of the respondents, as some might regard the unhealthy options as healthy. Moreover, it was tried to keep the taste and look of the meals more or less similar, such that personal preferences would not solely determine the choices of the participants. However, this could possibly lead to mixing up the healthy and unhealthy options as well. Also, there were only two options available, while in reality people generally have more food options to choose from. Subsequently, respondents had to make hypothetical choices. This could bias the results as people tend to make different choices in hypothetical situations compared to real life situations.

There are several solutions that might mitigate these issues in further research. It would be interesting to see whether the choices that were made would differ if the options are more divergent in for example healthiness, taste and size. Furthermore, more meal options should be made available, such that desire or aversion for specific ingredients would be ruled out as much as possible.

Some concerns can be raised regarding the effectiveness of the BTS in this study. The BTS method assumes that people use a common prior as reference point for predicting the distribution. As discussed in paragraph 5.1 of the results, the common prior assumption does not hold for the answers to question two. Even though it is unclear whether this reduces the effectiveness of the BTS, it indicates that telling the truth was possibly not utility-maximising behaviour for the participants. Also, it might be the case that the sample is not large enough to prevent an individual prediction or response from having a significant impact on the sample results.

Furthermore, it is assumed that the BTS decreases the level of social desirability bias (SDB), and that SDB causes discrepancies between the reported behaviour of the respondents and their actual behaviour. However, the level of SDB has not been measured. Therefore, it would be interesting to measure the level of SDB in further research, to see whether SDB is indeed the underlying factor causing these discrepancies.

Moreover, it could be that the respondents in this experiment did not feel the urge 'to lie' about their food intakes, thereby not being subject to SDB. A possibility for further research could be to test the effect of the BTS on people that are trying to lose weight without any successful results, as these people are more likely to be subject to SDB.

Finally, for simplicity reasons a one-shot experiment was used, which is not that realistic. Therefore, this study might lack external validity. In real life, food decisions have to be made constantly and can be interdependent. Some people might try to compensate a 'cheat day' by a period of healthy eating or might reward themselves with something sweet after a healthy eating day. This overall balance is difficult to measure in a one-shot experiment and might bias the results of this study. Hence, it is recommended to track the food intakes of the participants over a longer period of time in further research.

A way to tackle most concerns is performing the experiment of which the general setup was presented in the data chapter. Based on the knowledge gained in this study, this setup can be extended. The proposed experiment is a framed field experiment where subjects have to report their food intakes over a longer period of time, for example by using a food diary app. The control group would only have to track their own food intakes, whereas the treatment group would have to predict the distribution of the others' answers as well. As an incentive, participants could for example gain points which would give discount at gyms or grocery shops. People in the control group could gain points by consistently reporting their food intakes, whereas people in the treatment group would be rewarded for honest reporting. Within-subject effects could be measured by first letting the participants use a general food tracker app before adding the incentives scheme. Running a framed field experiment would make it possible to measure for example the weight of the participants. This makes it easier to measure the physical impact of the BTS on the dietary attempts of the participants.

7. CONCLUSION

This paper has examined the effectiveness of the Bayesian Truth Serum on dietary self-report measures and to what extent it incentivizes people to make healthier choices. Self-report measures, including food diaries, often do not give a representative view of the actual behaviour of the respective respondents. Social desirability bias could possibly explain why using a food diary has not the desired effect on the health of the users. This bias implies that people have the tendency to give socially desirable responses to present a favourable image of the self either to oneself or others, also known as self-deception and impression management respectively. Previous research shows that the Bayesian Truth Serum can reduce socially desirable responding. The Bayesian Truth Serum is a survey scoring method that can be used to elicit truthful subjective data, as respondents are incentivized to report honestly. Respondents do not only have to report their own behaviour, but have to predict the distribution of the answers of the other respondents as well. Truthfulness is rewarded by giving a higher score to the answers that are more common than collectively predicted.

A questionnaire is used to test whether being exposed to the Bayesian Truth Serum induces people to make healthier meal choices. First, respondents had to answer ten questions about their current food and drink intakes. These questions regarded both healthy and unhealthy behaviour, which could be either confirmed or denied by answering ‘Yes’ or ‘No’. Based on these answers, participants received a health score, that ranged from zero to ten points, depending on how often they gave the healthy answer. Subsequently, respondents had to make five hypothetical meal choices: one for breakfast, lunch, afternoon snack, dinner and dessert. There were two options the respondents could choose from: a healthy and unhealthy option. Again, a choice score was given of zero to five points, depending on how often they chose the healthy option. Respondents that were randomly selected to be in the control group only had to answer the questions for themselves, whereas subjects in the treatment group had to predict the general distributions of their own answers as well.

Fisher’s exact tests and Mann-Whitney U test were employed to analyse the data. Furthermore, parametric tests (OLS, binomial and ordered Probit regression models) were used to check for robustness of the nonparametric results.

Based on the literature, two hypotheses were formulated.

The first hypothesis is focused on the effect of the Bayesian Truth Serum on the answers about the respondents' current dietary intakes: *Participants that are exposed to the Bayesian Truth Serum will be more honest about their eating behaviour.*

The results of the Fisher's exact tests indicate that there is no significant difference in the answers to the questions about current intakes between the control and treatment group. Moreover, the Mann-Whitney U test showed that the average health scores of the control and treatment group did not differ significantly. The binomial Probit regression models that were employed for the healthy answers to the ten questions indicated that the Bayesian Truth Serum did not significantly affect the probability of giving the healthy answer, and the sign of the effect differed per question. The results of the OLS and ordered Probit regression models were insignificant as well. These results indicate that there is no support for the first hypothesis. Thus, there is no evidence that the Bayesian Truth Serum influenced the answers of the respondents, implying that the respondents in the treatment group were not more honest about their current intakes than the respondents in the control group.

This contradicts the expectations that were formed in the theoretical framework. As social desirability bias can cause more dishonest behaviour, and previous research showed that the Bayesian Truth Serum can reduce the number of socially desirable answers, it was expected that respondents in the treatment group would give different, i.e. less healthy, answers than the respondents in the control group. Therefore, the Bayesian Truth Serum does not have the desired effect of making people more honest about, and thereby hopefully more aware of, their eating behaviour.

The second hypothesis concerns the effect of the Bayesian Truth Serum on the meal choices the respondents had to make: *Participants that are exposed to the Bayesian Truth Serum will choose the healthy meal alternative more often than participants that are not incentivized to report honestly.*

The results of the Fisher's exact tests and Mann-Whitney U test indicate that respondents in the treatment group did not significantly make different meal choices than the respondents in the control group, nor that their choice scores significantly differed. The binomial Probit regression models for the five separate meal choices suggested that the Bayesian Truth Serum only significantly decreased the probability of choosing the healthy breakfast option with 37.6 percentage points, while no significant evidence was found for the effect of the Bayesian Truth Serum on the other hypothetical meal choices. Again, the effects of the Bayesian Truth Serum on the choice score and the probabilities of having a specific choice score were insignificant in the OLS and ordered Probit regression models respectively. Thus, no evidence is found that respondents in the treatment group chose the healthy meal options more often, meaning that the hypothesis cannot be supported.

According to the cognitive dissonance theory, it was expected that people would try to solve psychological discomfort by choosing the healthier option if they became aware of their, possibly unhealthy, behaviour. However, it could also be that the Bayesian Truth Serum reduced social desirability bias, thereby incentivizing people to choose the unhealthy option more often if this was the consequence of showing more honest behaviour. Hence, the effect of the Bayesian Truth Serum was regarded as ambiguous in this case. Nonetheless, as no significant evidence is found that the latter effect has occurred, it cannot be said with certainty that the Bayesian Truth Serum induces people to make less healthier choices.

To conclude, the following research question can be answered:

Do people make healthier choices if they are incentivized to be honest about their eating behaviour?

No evidence is found that the Bayesian Truth Serum effectively incentivizes respondents to be more honest about their food and drink intakes, nor that these respondents make healthier choices after being exposed to the Bayesian Truth Serum. Thus, people do not make healthier choices if they are incentivized to be honest about their eating behaviour.

This study has tried to shed some light on reporting eating behaviour with self-monitoring measures and the possible consequences for the health and dieting industry. The Bayesian Truth Serum appears to be no effective instrument for stimulating people to make healthier choices. Further research can rule out whether this is still the case for real-life eating decisions if the Bayesian Truth Serum is applied in a field setting.

8. BIBLIOGRAPHY

ANP. (2010, November 2). Overgewicht kost 3 miljard per jaar. *De volkskrant*.

Bachkirova, T. (2016). A new perspective on self-deception for applied purposes. *New Ideas in Psychology, 43*, 1-9.

Bandura, A. (1998). Health promotion from the perspective of social cognitive theory. *Psychology and Health, 13*(4), 623-649.

Barrage, L., & Lee, M. (2010). A penny for your thoughts: Inducing truth-telling in stated preference elicitation. *Economics letters, 106*(2), 140-142.

Bayes, T., Price, R., & Canton, J. (1763). *An essay towards solving a problem in the doctrine of chances.*

Bernardic, U. (2017). *The impact of oath and bayesian truth serum on self-deception.* Retrieved from UvA Scripties Online: http://www.scriptiesonline.uba.uva.nl/639715

Bernardo, J. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological), 41*(2), 113-147.

Bertrand, M., & Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. *The American Economic Review, 91*(2), 67-72.

Blokstra, A., Burns, C., & Seidell, J. (1999). Perception of weight status and dieting behaviour in Dutch men and women. *International Journal of Obesity, 23*(1), 7-18.

Bovell-Benjamin, A. (2007). Sweet Potato: A Review of its Past, Present, and Future Role in Human Nutrition. *Advances in Food and Nutrition Research, 52*, 1-59.

Brown, I., Tzoulaki, I., Candeias, V., & Elliot, P. (2009). Salt intakes around the world: implications for public health. *International Journal of Epidemiology, 38*(3), 791-813.

Burke, L., Wang, J., & Sevick, M. (2011). Self-monitoring in weight loss: a systematic review of the literature. *Journal of the American Dietic Association, 111*(1), 92-102.

Camire, M., Kubow, S., & Donnelly, D. (2009). Potatoes and human health. *Critical Reviews in Food Science and Nutrition, 49*, 823-840.

Carlak, A. (2018, March 2). *Meer fastfoodrestaurants in de stad; vooral pizza populair*. Retrieved from Het Parool: https://www.parool.nl/amsterdam/meer-fastfoodrestaurants-in-de-stad-vooral-pizza-populair~a4576449/

Centraal Bureau voor de Statistiek. (2018, April 6). *Overgewicht.* Retrieved from Volksgezondheidenzorg.info: https://www.volksgezondheidenzorg.info/onderwerp/overgewicht/cijfers-context/huidige-situatie#node-overgewicht-volwassenen

Cummings, R., Elliot, S., Harrison, G., & Murphy, J. (1997). Are hypothetical referenda incentive compatible? *Journal of Political Economy, 105*(3), 609-621.

Czopp, A., Monteith, M., & Mark, A. (2006). Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology, 90*(5), 784-803.

Dawes, R. (1989). Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology, 25*, 1-17.

Dawes, R. (1990). The potential non-falsity of the false consensus effect. In R. (. Hogarth, *Insights in decision making. A tribute to Hillel J. Einhorn* (pp. 179-199). London: The University of Chicago Press.

de Vries, H., Kooiman, T., van Ittersum, M., van Brussel, M., & de Groot, M. (2016). Do activity monitors increase physical activity in adults with overweight or obesity? A systematic review and meta-analysis. *Obesity, 24*(10), 2078-2091.

Doll, T., Fulgoni III, V., Zhang, Y., Reimers, K., Packard, P., & Astwood, J. (2009). Potential health benefits and medical cost savings from calorie, sodium, and saturated fat reductions in the American diet. *American Journal of Health Promotion, 23*(6), 412-422.

Dreher, M. (2012, April 1). Pistachio nuts: composition and potential health benefits. *Nutrition Reviews, 70*(4), 234-240.

Festinger, L. (1957). A theory of cognitive dissonance. In L. Festinger, *Cognitive dissonance.* Stanford: Stanford University Press.

Fisher, R. (1993). Social desirability bias and the validity of indirect questioning. *Journal of consumer research, 20*(2), 303-315.

Greger, M. (2016, August 5). *How Not To Die: The Role of Diet in Preventing, Arresting, and Reversing Our Top 15 Killers*. Retrieved from Nutritionfacts.org: https://nutritionfacts.org/video/how-not-to-die/

Hebert, J., Clemow, L., Pbert, L., Ockene, I., & Ockene, J. (1995). Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. *International Journal of Epidemiology, 24*(2), 389-398.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review, 91*(2), 73-78.

Hogan, R. (1983). A socioanalytic theory of personality. In M. Page (Ed.), *Nebraska symposium on motivation* (Vol. 29, pp. 55-89). Lincoln: University of Nebraska Press.

Huang, C., Liao, H., & Chang, S. (1998). Social desirability and the clinical self-report inventory: Methodological reconsideration. *Journal of Clinical Psychology, 54*(4), 517-528.

John, L., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science, 23*(5), 524-532.

King, M., & Bruner, G. (2000). Social desirability bias: a neglected aspect of validity testing. *Psychology and Marketing, 17*(2), 79-103.

Klesges, R., Eck, L., & Ray, J. (1995). Who underreports dietary intake in a dietary recall? Evidence from the second national health and nutrition examination survey. *Journal of consulting and clinical psychology, 63*(3), 438-444.

Lustig, R., Schmidt, L., & Brindis, C. (2012). Public health: the toxic truth about sugar. *Nature, 482*(7383), 27-29.

Maccoby, E., & Maccoby, N. (1954). The interview: A tool of social science. In L. Gardiner, *Handbook of social psychology* (Vol. 1, pp. 449-487). Cambridge, MA: Addison-Wesley.

Martocchio, J., & Judge, T. (1997). Relationship between conscientiousness and learning in employee training: Mediating influences of self-deception and self-efficacy. *Journal of Applied Psychology, 82*(5), 764-773.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45*(6), 633-644.

Muhlheim, L., Allison, D., Heshka, S., & Heymsfield, S. (1998). Do unsuccessful dieters intentionally underreport food intake? *International Journal of Eating Disorders, 24*(3), 259-266.

Navruz-Varli, S., & Sanlier, N. (2016). Nutritional and health benefits of quinoa (Chenopodium quioa Willd.). *Journal of Cereal Science, 69*, 371-376.

Nederhof, A. (1985). Methods of coping with social desirability bias: a review. *European Journal of Social Psychology, 15*(3), 263-280.

Netherlands Nutrition Centre. (2018). *Acrylamide*. Retrieved July 2018, from Voedingscentrum: https://www.voedingscentrum.nl/encyclopedie/acrylamide.aspx

Netherlands Nutrition Centre. (2018). *Bloemkool*. Retrieved July 2018, from Voedingscentrum: https://www.voedingscentrum.nl/encyclopedie/bloemkool.aspx

Netherlands Nutrition Centre. (2018). *Groente en fruit*. Retrieved July 2018, from Voedingscentrum: https://www.voedingscentrum.nl/nl/gezond-eten-met-de-schijf-van-vijf/wat-staat-er-in-de-vakken-van-de-schijf-van-vijf/groente-en-fruit.aspx

Netherlands Nutrition Centre. (2018). *Is havermout gezond?* Retrieved July 2018, from Voedingscentrum: https://www.voedingscentrum.nl/nl/service/vraag-en-antwoord/gezonde-voeding-en-voedingsstoffen/is-havermout-gezond.aspx

Netherlands Nutrition Centre. (2018). *Wat is een gezond ontbijt?* Retrieved July 2018, from Voedingscentrum: https://www.voedingscentrum.nl/nl/service/vraag-en-antwoord/gezonde-voeding-en-voedingsstoffen/wat-is-een-gezond-ontbijt-.aspx

Netherlands Nutrition Centre. (2018). *Wat zijn gezonde toetjes?* Retrieved July 2018, from Voedingscentrum: https://www.voedingscentrum.nl/nl/service/vraag-en-antwoord/gezonde-voeding-en-voedingsstoffen/wat-zijn-gezonde-toetjes-en-desserts.aspx

Olshausen, B. (2004, March 1). *Bayesian probability theory*. Retrieved from Redwood Center for Theoretical Neuroscience - UC Berkely: http://www.rctn.org/bruno/npb163/bayes.pdf

Paulhus, D. (1984). Two-component models of socially desirable responding. *Journal of personality and social psychology, 46*(3), 598-609.

Paulhus, D. (1991). Measurement and control of response bias. In J. Robinson, P. Shaver, & L. Wrightsman, *Measures of personality and social psychology attitudes* (pp. 17-59). New York: Harcourt Brace Jovanovich.

Paulhus, D., & Vazire, S. (2009). The self-report method. In R. Robins, R. Fraley, & R. Krueger, *Handbook of research methods in personality psychology* (pp. 224-239). New York: Guilford Press.

Paulhus, D., Harms, P., Bruce, M., & Lysy, D. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of personality and social psychology, 84*(4), 890-904.

Pel, A.-F. (2016, April 15). *Pas op met wat foodblogger je voorschotelt*. Retrieved from Metro Nieuws: https://www.metronieuws.nl/nieuws/binnenland/2016/04/pas-op-met-wat-foodblogger-je-voorschotelt

Prelec, D. (2004). A bayesian truth serum for subjective data. *Science, 306*(5695), 462-466.

Prelec, D., & Seung, S. (2006). An algorithm that finds truth even if most people are wrong. *MIT Working Paper*.

Roth, D., Snyder, C., & Pace, L. (1986). Dimensions of favorable self-presentation. *Journal of personality and social psychology, 51*(4), 867-874.

RTL Nieuws. (2016, Augustus 6). *Deskundigen waarschuwen: mensen te veel bezig met gezond eten*. Retrieved from RTL Nieuws: https://www.rtlnieuws.nl/gezondheid/deskundigen-waarschuwen-mensen-te-veel-bezig-met-gezond-eten

Sackeim, H., & Gur, R. (1979). Self-deception, other-deception, and self-reported psychopathology. *Journal of consulting and clinical psychology, 47*(1), 213-215.

Sackeim, H., & Gur, R. (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology, 37*(2), 147-169.

Sartre, J. (1958). Being and nothingness: An essay on phenomenological ontology. London: Methuen.

Spence, J., Jenkins, D., & Davignon, J. (2010). Dietary cholesterol and egg yolks: not for patients at risk of vascular disease. *Canadian Journal of Cardiology, 26*(9), e336-e339.

Spreadbury, I. (2012). Comparison with ancestral diets suggests dense acellular carbohydrates promote an inflammatory microbiota, and may be the primary dietary cause of leptin resistance and obesity. *Diabetes, metabolic syndrome and obesity: targets and therapy, 5*, 175-189.

Stice, E., Mazotti, L., Weibel, D., & Agras, W. (2000). Dissonance prevention program decreases thin-ideal internalization, body dissatisfaction, dieting, negative affect, and bulumic symptoms: A preliminary experiment. *International Journal of Eating Disorders, 27*(2), 206-217.

Surowiecki, J. (2004). The wisdom of Crowds. New York: Random House.

Taylor, S., & Brown, J. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological bulletin, 103*(2), 193-210.

Tirole, J. (2002). Rational irrationality: Some economics of self-management. *European Economic Review, 46*(4-5), 633-655.

Van der Zee, R. (2016, April 11). *Vergeet Sonja Bakker en Atkins: diëten maakt dikker*. Retrieved from HP de tijd: https://www.hpdetijd.nl/2016-04-11/vergeet-sonja-bakker-en-atkins-dieten-maakt-dikker/

Van Duyn, M., & Pivonka, E. (2000). Overview of the health benefits of fruit and vegetable consumption for the dietetics professional: selected literature. *Journal of the American Dietetic Assocation, 100*(12), 1511-1521.

Weaver, R., & Prelec, D. (2013). Creating truth-telling incentives with the bayesian truth serum. *Journal of Marketing Research, 50*(3), 289-302.

Weiss, R. (2009). *Optimally aggregating elicited expertise: a proposed application of the bayesian truth serum for policy analysis.* Retrieved from Doctoral dissertation, Massachusetts Institute of Technology: http://hdl.handle.net/1721.1/53066

9. APPENDIX

9.1 Description of the variables in the dataset

Description of the dataset 1: The variables

Variable	Explanation
Age	Age in years
Male	Indicator whether an individual is male (1) or female (0)
Nationality	Indicator whether the nationality of a subject is Belgium (0), Bulgaria (1), China (2), Lebanon (3), Netherlands (4) or UK (5); split into six categorical dummies
Education level	Indicator whether the highest education level of a subject is primary school (0), high school (1), college (2), bachelor degree (3), master degree (4) or doctorates degree (5); split into six categorical dummies
Vegan	Indicator whether an individual is vegan or vegetarian (1) or not (0)
BTS	Indicator whether an individual is in the BTS treatment group (1) or in the control group (0)
Q1-10	Answer to questions 1 to 10, yes (1) or no (0)
Breakfast_healthy	Indicator whether an individual chooses the healthy breakfast option (1) or the unhealthy alternative (0)
Lunch_healthy	Indicator whether an individual chooses the healthy lunch option (1) or the unhealthy alternative (0)
Snack_healthy	Indicator whether an individual chooses the healthy snack option (1) or the unhealthy alternative (0)
Dinner_healthy	Indicator whether an individual chooses the healthy dinner option (1) or the unhealthy alternative (0)
Dessert_healthy	Indicator whether an individual chooses the healthy dessert option (1) or the unhealthy alternative (0)
Health Score	Score on an interval scale from 0 to 10 that indicates how often the individual gives the healthy answer to question 1 to 10
Choice Score	Score on an interval scale of 0 to 5 that indicates how often the individual chooses the healthy meal alternative for breakfast, lunch, snack, dinner and dessert

Description of the dataset 2: Healthy and unhealthy answer/choice categories

Variable	Healthy answer	Unhealthy answer
Question 1 - breakfast	Yes	No
Question 2 – sugar	No	Yes
Question 3 – wholemeal	Yes	No
Question 4 – soft drinks	No	Yes
Question 5 – instant meals	No	Yes
Question 6 – fruit	Yes	No
Question 7 – fried food	No	Yes
Question 8 – vegetables	Yes	No
Question 9 – take-out	No	Yes
Question 10 – snacks	No	Yes
Breakfast	(soy) yogurt with oatmeal and fresh fruit	Pancakes with maple syrup and fruit
Lunch	Quinoa salad with roasted vegetables	Caesar salad with grilled chicken
Afternoon snack	Pistachio nuts	Chips
Dinner	Cauliflower steak with sweet potato fries	Steak with French fries
Dessert	Fresh strawberries	Strawberry cheesecake

9.2 Descriptive statistics

9.2.1 Demographics

Descriptive statistics 1: Mean Age

Variable	Obs.	Mean	Std. Dev	Min	Max
Age	132	36.303	16.185	13	78

Descriptive statistics 2: Frequencies Gender

Gender	Freq.	Percent	Cum.
Female	93	70.45	70.45
Male	39	29.55	100.00
Total	132	100.00	

Descriptive statistics 3: Frequencies Nationality

Nationality	Freq.	Percent	Cum.
Belgium	2	1.52	1.52
Bulgaria	1	0.76	2.27
China	1	0.76	3.03
Lebanon	1	0.76	3.79
Netherlands	126	95.45	99.24
UK	1	0.76	100.00
Total	132	100.00	

Descriptive statistics 4: Frequencies Education Level

Education level	**Freq.**	**Percent**	**Cum.**
Primary school	2	1.52	1.52
High school	23	17.42	18.94
College	18	13.64	32.58
Bachelor graduate	38	28.79	61.37
Master graduate	48	36.36	97.73
Doctorates degree	3	2.27	100.00
Total	132	100.00	

Descriptive statistics 5: Frequencies Vegan/Vegetarian

Vegan/Vegetarian	**Freq.**	**Percent**	**Cum.**
No	124	93.94	93.94
Yes	8	6.06	100.00
Total	132	100.00	

Descriptive statistics 6: Frequencies Treatment

Treatment	**Freq.**	**Percent**	**Cum.**
Control	67	50.76	50.75
BTS	65	49.24	100.00
Total	132	100.00	

9.2.2 Demographics per treatment group

Descriptive statistics 7: Mean Age per Treatment Group

	Control group		BTS treatment group	
Variable	**Mean**	**Std. Dev**	**Mean**	**Std. Dev**
Age	35.881	16.752	36.738	15.698

Descriptive statistics 8: Frequencies Gender per Treatment Group

	Control group		BTS treatment group	
Gender	**Freq.**	**Percent**	**Freq.**	**Percent**
Female	45	67.16	48	73.85
Male	22	32.84	17	26.15
Total	67	100.00	65	100.00

Descriptive statistics 9: Frequencies Nationality per Treatment Group

	Control group		BTS treatment group	
Nationality	Freq.	Percent	Freq.	Percent
Belgium	2	2.99	0	0.00
Bulgaria	1	1.49	0	0.00
China	0	0.00	1	1.54
Lebanon	0	0.00	1	1.54
Netherlands	64	95.52	62	95.38
UK	0	0.00	1	1.54
Total	67	100.00	65	100.00

Descriptive statistics 10: Frequencies Education Level per Treatment Group

	Control group		BTS treatment group	
Education level	Freq.	Percent	Freq.	Percent
Primary school	2	2.99	0	0.00
High school	11	16.42	12	18.46
College	10	14.93	8	12.31
Bachelor graduate	19	28.36	19	29.23
Master graduate	25	37.31	23	35.38
Doctorates degree	0	0.00	3	4.62
Total	67	100.00	65	100.00

Descriptive statistics 11: Frequencies Vegan/Vegetarian per Treatment Group

	Control group		BTS treatment group	
Vegan/Vegetarian	**Freq.**	**Percent**	**Freq.**	**Percent**
No	63	94.03	61	93.85
Yes	4	5.97	4	6.15
Total	67	100.00	65	100.00

9.2.3 The answers of the participants

Descriptive statistics 12: Frequencies of Healthy Answers per Treatment Group

	Control group		**BTS treatment group**	
Variable	Frequency of healthy answer	% of healthy answer	Frequency of healthy answer	% of healthy answer
Question 1	51	76.12	49	75.38
Question 2	64	95.52	61	93.85
Question 3	52	77.61	52	80.00
Question 4	59	88.06	59	90.77
Question 5	52	77.61	49	75.38
Question 6	33	49.25	32	49.23
Question 7	40	59.70	35	53.85
Question 8	49	73.13	41	63.08
Question 9	55	82.09	58	89.23
Question 10	35	52.24	27	41.54
Breakfast	57	85.07	54	83.08
Lunch	22	32.84	26	40.00
Afternoon snack	40	59.70	43	66.15
Dinner	27	40.30	27	41.54
Dessert	44	65.67	35	53.85
Observations	67		65	

9.2.4 Health Score and Choice Score

Descriptive statistics 13: Mean Health Score and Choice Score per Treatment Group

	Control group		**BTS treatment group**	
Variable	**Mean**	**Std. Dev**	**Mean**	**Std. Dev**
Health score	7.313	1.672	7.123	2.019
Choice score	2.836	1.136	2.846	1.240

9.3 Nonparametric tests

9.3.1 Analysis of the relation between BTS treatment and being honest about current dietary intakes

Result 1: Fisher's Exact test - Answer to question one vs. Treatment

	Treatment		
Answer to question 1	Control group	BTS group	Total
No	16	16	32
Yes	51	49	100
Total	67	65	132
	Fisher's exact	=	1.000

Result 2: Fisher's Exact test - Answer to question two vs. Treatment

	Treatment		
Answer to question 2	Control group	BTS group	Total
No	64	61	125
Yes	3	4	7
Total	67	65	132
	Fisher's exact	=	0.716

Result 3: Fisher's Exact test - Answer to question three vs. Treatment

	Treatment		
Answer to question 3	Control group	BTS group	Total
No	15	13	28
Yes	52	52	104
Total	67	65	132
	Fisher's exact	=	0.832

Result 4: Fisher's Exact test - Answer to question four vs. Treatment

	Treatment		
Answer to question 4	Control group	BTS group	Total
No	59	59	118
Yes	8	6	14
Total	67	65	132
	Fisher's exact	=	0.779

Result 5: Fisher's Exact test - Answer to question five vs. Treatment

	Treatment		
Answer to question 5	Control group	BTS group	Total
No	52	49	101
Yes	15	16	31
Total	67	65	132
	Fisher's exact	=	0.838

Result 6: Fisher's Exact test - Answer to question six vs. Treatment

	Treatment		
Answer to question 6	Control group	BTS group	Total
No	34	33	67
Yes	33	32	65
Total	67	65	132
	Fisher's exact	=	1.000

Result 7: Fisher's Exact test - Answer to question seven vs. Treatment

	Treatment		
Answer to question 7	Control group	BTS group	Total
No	40	35	75
Yes	27	30	57
Total	67	65	132
	Fisher's exact	=	0.598

Result 8: Fisher's Exact test - Answer to question eight vs. Treatment

	Treatment		
Answer to question 8	Control group	BTS group	Total
No	18	24	42
Yes	49	41	90
Total	67	65	132
	Fisher's exact	=	0.263

Result 9: Fisher's Exact test - Answer to question nine vs. Treatment

	Treatment		
Answer to question 9	Control group	BTS group	Total
No	55	58	113
Yes	12	7	19
Total	67	65	132
	Fisher's exact	=	0.323

Result 10: Fisher's Exact test - Answer to question ten vs. Treatment

	Treatment		
Answer to question 10	Control group	BTS group	Total
No	35	27	62
Yes	32	38	70
Total	67	65	132
	Fisher's exact	=	0.228

Result 11: Mann Whitney U Test - Health Score vs. Treatment

BTS	**Obs.**	**Rank sum**	**Expected**
0	67	4513.50	4455.50
1	65	4264.50	4322.50
Combined	132	8778	8778
	z	=	0.268
	Prob > |z|	=	0.7890

9.3.2 Analysis of the relation between BTS treatment and hypothetical meal choices

Result 12: Fisher's Exact Test - Breakfast Choice vs. Treatment

	Treatment		
Breakfast	Control group	BTS group	Total
Unhealthy	10	11	21
Healthy	57	54	111
Total	67	65	132
	Fisher's exact	=	0.815

Result 13: Fisher's Exact Test - Lunch Choice vs. Treatment

	Treatment		
Lunch	Control group	BTS group	Total
Unhealthy	45	39	84
Healthy	22	26	48
Total	67	65	132
	Fisher's exact	=	0.470

Result 14: Fisher's Exact Test - Snack Choice vs. Treatment

	Treatment		
Snack	Control group	BTS group	Total
Unhealthy	27	22	49
Healthy	40	43	83
Total	67	65	132
	Fisher's exact	=	0.475

Result 15: Fisher's Exact Test - Dinner Choice vs. Treatment

	Treatment		
Dinner	Control group	BTS group	Total
Unhealthy	40	38	78
Healthy	27	27	54
Total	67	65	132
	Fisher's exact	=	1.000

Result 16: Fisher's Exact Test - Dessert Choice vs. Treatment

	Treatment		
Dessert	Control group	BTS group	Total
Unhealthy	23	30	53
Healthy	44	35	79
Total	67	65	132
	Fisher's exact	=	0.214

Result 17: Mann Whitney U Test - Choice Score vs. Treatment

BTS	**Obs.**	**Rank sum**	**Expected**
0	67	4422	4455.5
1	65	4356	4322.50
Combined	132	8778	8778
	z	=	-0.157
	Prob > |z|	=	0.8752

9.4 Parametric tests

9.4.1 Analysis of the relation between BTS treatment and being honest about current dietary intakes

Result 18: Probit Regression Model – Healthy Answer to Question One

Question 1 (Yes=1)	**Coefficient**	**Std. Err.**	**p-value**
BTS	-0.090	0.250	0.720
Age	0.026	0.009	0.003***
Male	-0.300	0.263	0.255
Constant	-0.025	0.343	0.941
Pseudo R^2	0.083		
Log likelihood	-67.041		
Observations	132		
Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.			

Result 19: Average Marginal Effects Analysis – Healthy Answer to Question One

Question 1 (Yes=1)	**dy/dx**	**Std. Err.**	**p-value**
BTS	-0.025	0.071	0.720
Age	0.007	0.002	0.001***
Male	-0.085	0.074	0.247
Observations	132		
Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.			

Result 20: Probit Regression Model – Healthy Answer to Question Two

Question 2 (No=1)	**Coefficient**	**Std. Err.**	**p-value**
BTS	-0.167	0.367	0.649
Age	0.003	0.012	0.816
Male	0.033	0.401	0.935
Constant	1.597	0.492	0.001***
Pseudo R^2	0.005		
Log likelihood	-27.247		
Observations	132		
****Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.***			

Result 21: Average Marginal Effects Analysis – Healthy Answer to Question Two

Question 2 (No=1)	**dy/dx**	**Std. Err.**	**p-value**
BTS	-0.018	0.040	0.650
Age	0.000	0.001	0.816
Male	0.004	0.043	0.935
Observations	132		
****Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.***			

Result 22: Probit Regression Model – Healthy Answer to Question Three

Question 3 (Yes=1)	**Coefficient**	**Std. Err.**	**p-value**
BTS	0.098	0.247	0.692
Age	0.004	0.008	0.610
Male	0.337	0.289	0.243
Constant	0.516	0.351	0.141
Pseudo R^2	0.012		
Log likelihood	-67.412		
Observations	132		
****Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.***			

Result 23: Average Marginal Effects Analysis – Healthy Answer to Question Three

Question 3 (Yes=1)	dy/dx	Std. Err.	p-value
BTS	0.028	0.071	0.692
Age	0.001	0.002	0.609
Male	0.097	0.082	0.238
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 24: Probit Regression Model – Healthy Answer to Question Four

Question 4 (No=1)	Coefficient	Std. Err.	p-value
BTS	0.093	0.321	0.772
Age	-0.013	0.010	0.197
Male	-1.135	0.329	0.001***
Constant	2.179	0.492	0.000***
Pseudo R^2	0.153		
Log likelihood	-37.835		
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 25: Average Marginal Effects Analysis – Healthy Answer to Question Four

Question 4 (Yes=1)	dy/dx	Std. Err.	p-value
BTS	0.015	0.050	0.772
Age	-0.002	0.002	0.198
Male	-0.178	0.051	0.000***
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 26: Probit Regression Model – Healthy Answer to Question Five

Question five (No=1)	Coefficient	Std. Err.	p-value
BTS	-0.119	0.253	0.639
Age	0.025	0.009	0.004***
Male	-0.451	0.265	0.089*
Constant	0.091	0.354	0.797
Pseudo R^2	0.097		
Log likelihood	-65.003		
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 27: Average Marginal Effects Analysis – Healthy Answer to Question Five

Question 5 (No=1)	dy/dx	Std. Err.	p-value
BTS	-0.033	0.070	0.638
Age	0.007	0.002	0.002***
Male	-0.125	0.071	0.079*
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 28: Probit Regression Model – Healthy Answer to Question Six

Question 6 (Yes=1)	Coefficient	Std. Err.	p-value
BTS	-0.030	0.222	0.893
Age	0.015	0.007	0.032**
Male	-0.212	0.247	0.390
Constant	-0.488	0.316	0.122
Pseudo R^2	0.034		
Log likelihood	-88.392		
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 29: Average Marginal Effects Analysis – Healthy Answer to Question Six

Question 6 (Yes=1)	**dy/dx**	**Std. Err.**	**p-value**
BTS	-0.011	0.085	0.893
Age	0.006	0.003	0.023**
Male	-0.082	0.094	0.386
Observations	132		
Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.			

Result 30: Probit Regression Model – Healthy Answer to Question Seven

Question 7 (No=1)	**Coefficient**	**Std. Err.**	**p-value**
BTS	-0.248	0.232	0.285
Age	0.020	0.007	0.007***
Male	-0.807	0.256	0.002***
Constant	-0.168	0.320	0.599
Pseudo R^2	0.116		
Log likelihood	-79.839		
Observations	132		
Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.			

Result 31: Average Marginal Effects Analysis – Healthy Answer to Question Seven

Question 7 (No=1)	**dy/dx**	**Std. Err.**	**p-value**
BTS	-0.085	0.079	0.279
Age	0.007	0.002	0.004***
Male	-0.278	0.078	0.000***
Observations	132		
Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.			

Result 32: Probit Regression Model – Healthy Answer to Question Eight

Question 8 (Yes=1)	Coefficient	Std. Err.	p-value
BTS	-0.282	0.229	0.218
Age	0.001	0.007	0.922
Male	0.026	0.254	0.919
Constant	0.583	0.326	0.074*
Pseudo R^2	0.009		
Log likelihood	-81.785		
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 33: Average Marginal Effects Analysis – Healthy Answer to Question Eight

Question 8 (Yes=1)	dy/dx	Std. Err.	p-value
BTS	-0.100	0.079	0.210
Age	0.000	0.003	0.922
Male	0.009	0.090	0.919
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 34: Probit Regression Model – Healthy Answer to Question Nine

Question 9 (No=1)	Coefficient	Std. Err.	p-value
BTS	0.341	0.307	0.266
Age	0.022	0.011	0.046**
Male	-1.025	0.301	0.001***
Constant	0.641	0.419	0.126
Pseudo R^2	0.187		
Log likelihood	-44.202		
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 35: Average Marginal Effects Analysis – Healthy Answer to Question Nine

Question 9 (No=1)	dy/dx	Std. Err.	p-value
BTS	0.062	0.055	0.260
Age	0.004	0.002	0.040**
Male	-0.187	0.050	0.000***
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 36: Probit Regression Model – Healthy Answer to Question Ten

Question 10 (No=1)	Coefficient	Std. Err.	p-value
BTS	-0.256	0.221	0.246
Age	-0.008	0.007	0.255
Male	0.181	0.244	0.459
Constant	0.279	0.313	0.373
Pseudo R^2	0.020		
Log likelihood	-89.402		
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 37: Average Marginal Effects Analysis – Healthy Answer to Question Ten

Question 10 (No=1)	dy/dx	Std. Err.	p-value
BTS	-0.100	0.084	0.238
Age	-0.003	0.003	0.248
Male	0.070	0.094	0.456
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 38: OLS Regression Model - Health Score

Health score	**Coeff.**	**Std. Err.**	**p-value**
BTS	-0.231	0.307	0.453
Vegan	1.342	0.670	0.048**
Age	0.032	0.010	0.003***
Male	-0.687	0.358	0.057*
Nationality Belgium	4.363	2.118	0.042**
Nationality Bulgaria	1.205	2.537	0.636
Nationality China	5.026	2.434	0.041**
Nationality Netherlands	3.002	1.736	0.086*
Nationality Lebanon	-2.668	2.686	0.323
High school	1.239	1.279	0.335
College	0.560	1.302	0.668
Bachelor	0.872	1.250	0.486
Master	0.618	1.271	0.628
Doctorates	1.618	1.750	0.357
Constant	2.537	2.099	0.229
Number of observations	132		
R^2	0.2577		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 39: Ordered Probit Regression Model - Health Score

Health score	Coefficient	Std. Err.	p-value
BTS	-0.146	0.187	0.437
Vegan	0.856	0.516	0.040**
Age	0.021	0.007	0.002***
Male	-0.453	0.220	0.039**
Nationality Belgium	6.930	195.046	0.972
Nationality Bulgaria	4.745	195.048	0.981
Nationality China	7.126	195.048	0.971
Nationality Netherlands	5.905	195.045	0.976
Nationality Lebanon	-11.021	776.879	0.989
High school	0.822	0.772	0.287
College	0.317	0.784	0.686
Bachelor	0.564	0.752	0.453
Master	0.365	0.766	0.634
Doctorates	0.925	1.060	0.383
Pseudo R^2	0.084		
Log likelihood	-236.913		
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 40: Average Marginal Effects Analysis - Health Score

	Health score	dy/dx	Std. Err.	p-value
BTS	2	$6.29e^{-9}$	$5.75e^{-6}$	0.999
	3	0.010	0.013	0.455
	4	0.008	0.011	0.457
	5	0.012	0.016	0.446
	6	0.017	0.023	0.442
	7	0.005	0.007	0.466
	8	-0.008	0.011	0.449
	9	-0.021	0.027	0.439
	10	-0.022	0.029	0.442
Vegan	2	$-3.70e^{-8}$	0.000	0.999
	3	-0.056	0.035	0.107

	4	-0.047	0.030	0.116
	5	-0.071	0.040	0.075*
	6	-0.102	0.051	0.048**
	7	-0.028	0.018	0.123
	8	0.048	0.029	0.102
	9	0.125	0.062	0.044**
	10	0.131	0.066	0.046**
Age	2	$-8.93e^{-10}$	$8.17e^{-7}$	0.999
	3	-0.001	0.001	0.046**
	4	-0.001	0.001	0.051*
	5	-0.002	0.001	0.016**
	6	-0.002	0.001	0.003***
	7	-0.001	0.000	0.075*
	8	0.001	0.001	0.024**
	9	0.003	0.001	0.002***
	10	0.003	0.001	0.006***
Male	2	$1.96e^{-8}$	0.000	0.999
	3	0.030	0.019	0.118
	4	0.025	0.015	0.101
	5	0.038	0.020	0.057*
	6	0.054	0.027	0.046**
	7	0.015	0.011	0.168
	8	-0.025	0.014	0.067*
	9	-0.066	0.033	0.047**
	10	-0.069	0.036	0.054*
Nationality Belgium	2	$-2.99e^{-7}$	0.000	0.999
	3	-0.453	12.738	0.972
	4	-0.378	10.639	0.972
	5	-0.576	16.223	0.972
	6	-0.824	23.207	0.972
	7	-0.226	6.373	0.972
	8	0.385	10.839	0.972
	9	1.013	28.503	0.972
	10	1.060	29.835	0.972
Nationality Bulgaria	2	$-2.05e^{-7}$	0.000	0.999
	3	-0.310	12.737	0.981

	4	-0.259	10.638	0.981
	5	-0.395	16.223	0.981
	6	-0.565	23.207	0.981
	7	-0.155	6.372	0.981
	8	0.264	10.839	0.981
	9	0.693	28.503	0.981
	10	0.726	29.835	0.981
Nationality China	2	$-3.08e^{-7}$	0.000	0.999
	3	-0.465	12.738	0.971
	4	-0.389	10.639	0.971
	5	-0.593	16.223	0.971
	6	-0.848	23.207	0.971
	7	-0.233	6.373	0.971
	8	0.396	10.840	0.971
	9	1.041	28.503	0.971
	10	1.090	29.835	0.971
Nationality Netherlands	2	$-2.55e^{-7}$	0.000	0.999
	3	-0.386	12.737	0.976
	4	-0.322	10.638	0.976
	5	-0.491	16.223	0.976
	6	-0.703	23.207	0.976
	7	-0.193	6.373	0.976
	8	0.328	10.839	0.976
	9	0.863	28.503	0.976
	10	0.903	29.834	0.976
Nationality Lebanon	2	$4.76e^{-7}$	0.000	0.999
	3	0.720	50.733	0.989
	4	0.601	42.370	0.989
	5	0.917	64.614	0.989
	6	1.311	92.433	0.989
	7	0.360	25.380	0.989
	8	-0.612	43.171	0.989
	9	-1.610	113.527	0.989
	10	-1.686	118.832	0.989
High School	2	$-3.55e^{-8}$	0.000	0.999
	3	-0.537	0.055	0.331

	4	-0.045	0.045	0.316
	5	-0.068	0.066	0.297
	6	-0.098	0.093	0.294
	7	-0.027	0.030	0.363
	8	0.046	0.044	0.304
	9	0.120	0.114	0.293
	10	0.126	0.120	0.296
College	2	$-1.37e^{-8}$	0.000	0.999
	3	-0.021	0.052	0.691
	4	-0.017	0.043	0.688
	5	-0.026	0.065	0.687
	6	-0.038	0.094	0.687
	7	-0.010	0.027	0.687
	8	0.018	0.044	0.687
	9	0.046	0.115	0.687
	10	0.048	0.120	0.687
Bachelor	2	$-2.44e^{-8}$	0.000	0.999
	3	-0.037	0.051	0.474
	4	-0.031	0.042	0.465
	5	-0.047	0.063	0.457
	6	-0.067	0.090	0.457
	7	-0.018	0.027	0.494
	8	0.031	0.042	0.459
	9	0.082	0.111	0.457
	10	0.086	0.116	0.458
Master	2	$-1.57e^{-8}$	0.000	0.999
	3	-0.024	0.051	0.641
	4	-0.020	0.042	0.636
	5	-0.030	0.064	0.634
	6	-0.043	0.092	0.635
	7	-0.012	0.026	0.649
	8	0.020	0.043	0.635
	9	0.053	0.112	0.635
	10	0.056	0.118	0.636
Doctorates	2	$-3.99e^{-8}$	0.000	0.999
	3	-0.060	0.073	0.410

	4	-0.050	0.061	0.404
	5	-0.077	0.090	0.392
	6	-0.110	0.127	0.388
	7	-0.030	0.038	0.428
	8	0.051	0.061	0.400
	9	0.135	0.155	0.384
	10	0.142	0.164	0.388
Observations	132			

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 41. Predictive margins of BTS on the Health score

Health score	BTS	Margin	Std. Err.	p-value
1	0	$4.21e^{-28}$	1.27^{e-24}	1.000
1	1	0.015	0.000	0.000***
2	0	0.031	0.016	0.054*
2	1	0.048	0.016	0.003*
3	0	0.036	0.016	0.028**
3	1	0.037	0.017	0.028**
4	0	0.073	0.023	0.002***
4	1	0.073	0.023	0.002***
5	0	0.182	0.035	0.000***
5	1	0.179	0.035	0.000***
6	0	0.187	0.034	0.000***
6	1	0.181	0.033	0.000***
7	0	0.206	0.036	0.000***
7	1	0.198	0.035	0.000***
8	0	0.185	0.036	0.000***
8	1	0.176	0.035	0.000***
9	0	0.100	0.029	0.001***
9	1	0.093	0.027	0.001***

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

9.4.2 Analysis of the relation between BTS treatment and hypothetical meal choices

Result 42: Probit Regression Model – Healthy Breakfast Choice

Breakfast (Healthy=1)	Coefficient	Std. Err.	p-value
BTS	-2.018	1.142	0.077*
Health score	0.013	0.120	0.912
BTS*Health score	0.296	0.167	0.076*
Age	0.045	0.014	0.002***
Male	-0.006	0.323	0.986
Constant	-0.411	0.933	0.659
Pseudo R^2	0.232		
Log likelihood	-44.419		
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 43: Average Marginal Effects Analysis – Healthy Breakfast Choice

Breakfast (Healthy=1)	dy/dx	Std. Err.	p-value
BTS	-0.376	0.204	0.065*
Health score	0.002	0.022	0.912
BTS*Health score	0.055	0.030	0.064*
Age	0.008	0.002	0.001***
Male	-0.001	0.060	0.986
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 44: Probit Regression Model – Healthy Lunch Choice

Lunch (Healthy=1)	Coefficient	Std. Err.	p-value
BTS	-0.111	-0.968	0.909
Health score	-0.035	0.104	0.739
BTS*Health score	0.035	0.130	0.0789
Age	-0.005	0.008	0.525
Male	-0.929	0.292	0.001***
Constant	0.247	0.772	0.749
Pseudo R^2	0.072		
Log likelihood	-80.256		
Observations	132		
Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.			

Result 45: Average Marginal Effects Analysis – Healthy Lunch Choice

Lunch (Healthy=1)	dy/dx	Std. Err.	p-value
BTS	-0.038	0.336	0.909
Health score	-0.012	0.036	0.739
BTS*Health score	0.012	0.045	0.789
Age	-0.002	0.003	0.523
Male	-0.322	0.090	0.000***
Observations	132		
Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.			

Result 46: Probit Regression Model – Healthy Snack Choice

Snack (Healthy=1)	Coefficient	Std. Err.	p-value
BTS	-0.905	1.038	0.383
Health score	0.180	0.108	0.095*
BTS*Health score	0.167	0.144	0.247
Age	0.026	0.008	0.002***
Male	0.388	0.273	0.155
Constant	-2.088	0.814	0.010***
Pseudo R^2	0.188		
Log likelihood	-70.669		
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 47: Average Marginal Effects Analysis – Healthy Snack Choice

Snack (Healthy=1)	dy/dx	Std. Err.	p-value
BTS	-0.274	0.311	0.379
Health score	0.054	0.032	0.085*
BTS*Health score	0.050	0.043	0.240
Age	0.008	0.002	0.000***
Male	0.117	0.081	0.148
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 48: Probit Regression Model – Healthy Dinner Choice

Dinner (Healthy=1)	Coefficient	Std. Err.	p-value
BTS	1.053	1.014	0.299
Health score	0.243	0.106	0.022**
BTS*Health score	-0.144	0.135	0.284
Age	-0.005	0.008	0.496
Male	-0.845	0.282	0.003***
Constant	-1.595	0.787	0.043**
Pseudo R^2	0.109		
Log likelihood	-79.534		
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 49: Average Marginal Effects Analysis – Healthy Dinner Choice

Dinner (Healthy=1)	dy/dx	Std. Err.	p-value
BTS	0.363	0.346	0.294
Health score	0.084	0.034	0.015**
BTS*Health score	-0.050	0.046	0.278
Age	-0.002	0.003	0.494
Male	-0.291	0.088	0.001***
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 50: Probit Regression Model – Healthy Dessert Choice

Dessert (Healthy=1)	Coefficient	Std. Err.	p-value
BTS	-0.494	0.937	0.598
Health score	0.048	0.100	0.632
BTS*Health score	0.024	0.127	0.848
Age	0.008	0.007	0.271
Male	-0.132	0.256	0.606
Constant	-0.187	0.741	0.801
Pseudo R^2	0.032		
Log likelihood	-86.064		
Observations	132		
Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.			

Result 51: Average Marginal Effects Analysis – Healthy Dessert Choice

Dessert (Healthy=1)	dy/dx	Std. Err.	p-value
BTS	-0.184	0.348	0.597
Health score	0.018	0.037	0.631
BTS*Health score	0.009	0.047	0.848
Age	0.003	0.003	0.264
Male	-0.049	0.095	0.605
Observations	132		
Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.			

Result 52: OLS Regression Model - Choice Score

Choice score	**Coeff.**	**Std. Err.**	**p-value**
BTS	-0.466	0.784	0.554
Health score	0.166	0.080	0.042**
BTS*Health score	0.060	0.104	0.567
Vegan	0.770	0.410	0.063*
Age	0.012	0.007	0.066*
Male	-0.476	0.218	0.031**
Nationality Belgium	-0.396	1.306	0.762
Nationality Bulgaria	2.569	1.553	0.101
Nationality China	1.267	1.512	0.404
Nationality Netherlands	0.428	1.075	0.691
Nationality Lebanon	0.425	1.622	0.794
High school	0.270	0.776	0.728
College	0.493	0.785	0.531
Bachelor	0.462	0.754	0.541
Master	0.571	0.766	0.458
Doctorates	1.117	1.054	0.292
Constant	0.403	1.380	0.771
Number of observations	132		
R^2	0.3603		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 53: Ordered Probit Regression Model - Choice Score

Choice score	**Coefficient**	**Std. Err.**	**p-value**
BTS	-0.637	0.834	0.445
Health score	0.188	0.085	0.027**
BTS*Health score	0.080	0.111	0.470
Vegan	0.890	0.448	0.047**
Age	0.013	0.007	0.064*
Male	-0.534	0.231	0.021**
Nationality Belgium	-0.253	1.489	0.865
Nationality Bulgaria	6.922	122.784	0.955
Nationality China	1.525	1.677	0.363
Nationality Netherlands	0.690	1.257	0.583
Nationality Lebanon	0.976	1.793	0.586
High school	0.306	0.831	0.713
College	0.612	0.808	0.491
Bachelor	0.612	0.840	0.466
Master	0.699	0.822	0.395
Doctorates	1.212	1.119	0.278
Pseudo R^2	0.151		
Log likelihood	-174.598		
Observations	132		

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 54: Average Marginal Effects Analysis - Choice Score

	Choice score	**dy/dx**	**Std. Err.**	**p-value**
BTS	0	0.008	0.012	0.526
	1	0.106	0.139	0.446
	2	0.078	0.103	0.451
	3	-0.018	0.026	0.501
	4	-0.100	0.131	0.445
	5	-0.074	0.099	0.453
Health score	0	-0.002	0.002	0.338
	1	-0.031	0.015	0.033**

	2	-0.023	0.010	0.029**
	3	0.005	0.004	0.226
	4	0.030	0.014	0.029**
	5	0.022	0.011	0.044**
BTS*Health score	0	-0.001	0.002	0.542
	1	-0.013	0.019	0.472
	2	-0.010	0.014	0.475
	3	0.002	0.003	0.522
	4	0.013	0.017	0.471
	5	0.009	0.013	0.477
Vegan	0	-0.011	0.011	0.347
	1	-0.148	0.079	0.059*
	2	-0.109	0.058	0.059*
	3	0.025	0.025	0.317
	4	0.140	0.072	0.052*
	5	0.103	0.053	0.052*
Age	0	-0.000	0.000	0.355
	1	-0.002	0.001	0.071*
	2	-0.002	0.001	0.069*
	3	0.000	0.000	0.274
	4	0.002	0.001	0.063*
	5	0.001	0.001	0.084*
Male	0	0.006	0.007	0.335
	1	0.089	0.040	0.026**
	2	0.065	0.028	0.022**
	3	-0.015	0.012	0.205
	4	-0.084	0.037	0.024**
	5	-0.062	0.030	0.038**
Nationality Belgium	0	0.003	0.018	0.868
	1	0.042	0.249	0.865
	2	0.031	0.181	0.865
	3	-0.007	0.041	0.866
	4	-0.040	0.234	0.865
	5	-0.029	0.173	0.865
Nationality Bulgaria	0	-0.083	1.47	0.955
	1	-1.155	20.484	0.955

	2	-0.845	14.996	0.955
	3	0.191	3.394	0.955
	4	1.087	19.284	0.955
	5	0.805	14.276	0.955
Nationality China	0	-0.018	0.026	0.488
	1	-0.254	0.279	0.362
	2	-0.186	0.211	0.378
	3	0.042	0.057	0.458
	4	0.240	0.267	0.369
	5	0.177	0.197	0.368
Nationality Netherlands	0	-0.008	0.017	0.624
	1	-0.115	0.208	0.580
	2	-0.084	0.157	0.591
	3	0.019	0.037	0.606
	4	0.108	0.199	0.586
	5	0.080	0.147	0.585
Nationality Lebanon	0	-0.012	0.024	0.623
	1	-0.163	0.298	0.584
	2	-0.119	0.223	0.593
	3	0.027	0.052	0.601
	4	0.153	0.284	0.589
	5	0.113	0.211	0.590
High school	0	-0.004	0.011	0.729
	1	-0.051	0.139	0.713
	2	-0.037	0.101	0.713
	3	0.008	0.023	0.716
	4	0.048	0.131	0.714
	5	0.036	0.097	0.714
College	0	-0.007	0.012	0.548
	1	-0.102	0.140	0.467
	2	-0.075	0.103	0.467
	3	0.017	0.025	0.498
	4	0.096	0.133	0.468
	5	0.071	0.099	0.473
Bachelor	0	-0.007	0.012	0.563
	1	-0.093	0.135	0.491

	2	-0.068	0.099	0.492
	3	0.015	0.024	0.519
	4	0.087	0.127	0.492
	5	0.065	0.095	0.497
Master	0	-0.008	0.013	0.507
	1	-0.112	0.137	0.396
	2	-0.009	0.101	0.396
	3	0.019	0.025	0.444
	4	0.110	0.130	0.398
	5	0.081	0.097	0.403
Doctorates	0	-0.014	0.019	0.448
	1	-0.202	0.188	0.283
	2	-0.148	0.138	0.282
	3	0.033	0.040	0.399
	4	0.190	0.177	0.281
	5	0.141	0.132	0.287
Observations	132			

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

Result 55. Predictive margins of BTS on the Choice score

Choice score	**BTS**	**Margin**	**Std. Err.**	**p-value**
0	0	0.004	0.005	0.471
0	1	0.008	0.008	0.347
1	0	0.137	0.033	0.000***
1	1	0.149	0.033	0.000***
2	0	0.258	0.039	0.000***
2	1	0.242	0.017	0.000***
3	0	0.304	0.041	0.000***
3	1	0.289	0.040	0.000***
4	0	0.214	0.037	0.000***
4	1	0.226	0.037	0.000***
5	0	0.084	0.024	0.000***
5	1	0.086	0.027	0.001***

Significant at 10 percent level. **Significant at 5 percent level. ***Significant at 1 percent level.

9.5 The questionnaire

9.5.1 BTS treatment group

Dear responder,

Thank you for participating!

I am currently investigating people's eating behaviour and the dietary choices they make. This survey will not take you more than **10 minutes** to answer. One respondent will be randomly chosen and win **€20**.

I would like to notify you that your answers are anonymous and will be treated carefully. Your data will stored only for the purpose of this research. If you wish to get the chance of winning the prize, you can fill in your email address at the end of the questionnaire in order to contact you in the future. Your email address will be used solely for contacting you about the prize.
However, if you wish to remain **completely anonymous**, it is also possible. In that case, you will not be able to win the prize.

For each complete answer, you will earn lottery tickets based on that answer's '**Truth Score**'. Truth scoring, recently invented by a MIT professor and published in the academic journal Science (Prelec, 2004), rewards you for answering truthfully. Your truth score is based on both your personal answers to each question, as well as your predictions about the answers of others. Even though only you know if you really answered truthfully or not, **people who tell the truth score higher overall**.

You are most likely to maximize the number of lottery tickets you get if you answer every item truthfully. By 'truthfully' I mean: consider each item carefully, answer honestly, and take care to avoid mistakes.

Just by participating you already have the chance of winning the prize! For every truthful answer to the questions, you will increase the chance to win. At the end, one lottery ticket will be randomly selected to be paid out for real.

First, you will get 10 questions about your current food and drink intakes.
Please choose the answer you identify most with and answer these questions **truthfully**.

Let the survey begin!

Do you eat breakfast every day?

- ○ Yes
- ○ No

Think about the other respondents answering this survey.

What percentage of them do you feel will provide the same answer as you in the previous question?

	0	10	20	30	40	50	60	70	80	90	100
Percentage of people with the same response as you											

Do you use sugar in your coffee or tea?

- ○ Yes
- ○ No

Think about the other respondents answering this survey.

What percentage of them do you feel will provide the same answer as you in the previous question?

	0	10	20	30	40	50	60	70	80	90	100
Percentage of people with the same response as you											

Do you eat whole grain bread, rice and pasta instead of white bread, rice and pasta?

- Yes
- No

Think about the other respondents answering this survey.

What percentage of them do you feel will provide the same answer as you in the previous question?

	0	10	20	30	40	50	60	70	80	90	100
Percentage of people with the same response as you											

Do you consume soft drinks (e.g. coca cola, Fanta, energy drinks) every day?

- Yes
- No

Think about the other respondents answering this survey.

What percentage of them do you feel will provide the same answer as you in the previous question?

	0	10	20	30	40	50	60	70	80	90	100
Percentage of people with the same response as you											

Do you eat instant meals at least once a week?

- Yes
- No

Think about the other respondents answering this survey.

What percentage of them do you feel will provide the same answer as you in the previous question?

	0	10	20	30	40	50	60	70	80	90	100
Percentage of people with the same response as you											

Do you eat at least two pieces of fruit per day?

- Yes
- No

Think about the other respondents answering this survey.

What percentage of them do you feel will provide the same answer as you in the previous question?

	0	10	20	30	40	50	60	70	80	90	100
Percentage of people with the same response as you											

Do you eat fried foods (e.g. chips, fries, croquettes) at least once a week?

- ○ Yes
- ○ No

Think about the other respondents answering this survey.

What percentage of them do you feel will provide the same answer as you in the previous question?

	0	10	20	30	40	50	60	70	80	90	100
Percentage of people with the same response as you											

Do you eat at least 200 grams of vegetables per day?

- ○ Yes
- ○ No

Think about the other respondents answering this survey.

What percentage of them do you feel will provide the same answer as you in the previous question?

	0	10	20	30	40	50	60	70	80	90	100
Percentage of people with the same response as you											

Do you order take-out at least once a week?

- Yes
- No

Think about the other respondents answering this survey.

What percentage of them do you feel will provide the same answer as you in the previous question?

	0	10	20	30	40	50	60	70	80	90	100
Percentage of people with the same response as you											

Do you eat at least one less healthy snack (e.g. cookies, chocolate, chips) per day?

- Yes
- No

Think about the other respondents answering this survey. What percentage of them do you feel will provide the same answer as you in the previous question?

	0	10	20	30	40	50	60	70	80	90	100
Percentage of people with the same response as you											

You are halfway there!

Imagine that you are doing groceries for the next day and you have to determine what you will eat.
For every meal you have the choice between two alternatives. Both alternatives are equally priced, available in the same supermarket and have the same preparation time.

Let's choose!

For breakfast, I prefer:

(Soy) Yogurt with oatmeal and fresh fruit

Pancakes with maple syrup and fresh fruit

Think about the other respondents answering this survey. What percentage of them do you feel will provide the same answer as you in the previous question?

0 10 20 30 40 50 60 70 80 90 100

Percentage of people with the same response as you

For lunch, I prefer:

Quinoa salad with roasted vegetables

Caesar salad with grilled chicken

Think about the other respondents answering this survey. What percentage of them do you feel will provide the same answer as you in the previous question?

0 10 20 30 40 50 60 70 80 90 100

Percentage of people with the same response as you

As an afternoon snack, I prefer:

A handful of potato chips

A handful of pistachio nuts

Think about the other respondents answering this survey. What percentage of them do you feel will provide the same answer as you in the previous question?

0 10 20 30 40 50 60 70 80 90 100

Percentage of people with the same response as you

For dinner, I prefer:

Cauliflower steak with sweet potato fries

Steak with French fries

○ ○

Think about the other respondents answering this survey. What percentage of them do you feel will provide the same answer as you in the previous question?

0 10 20 30 40 50 60 70 80 90 100

Percentage of people with the same response as you

As dessert, I prefer:

Strawberry cheesecake

Fresh strawberries

○ ○

Think about the other respondents answering this survey. What percentage of them do you feel will provide the same answer as you in the previous question?

0 10 20 30 40 50 60 70 80 90 100

Percentage of people with the same response as you

You are almost done! To finish the survey, please fill out the demographic questions below.

Are you vegetarian/ vegan?

- ○ Yes
- ○ No

What is your age?

[ ]

What is your gender?

- ○ Male
- ○ Female

What is your nationality?

[▾]

What is the highest level of education you have completed?

- ○ High school
- ○ College
- ○ Bachelor graduate (university)
- ○ Master graduate (university)
- ○ Doctorates degree
- ○ [ ] Other, namely

Please enter your email address if you wish to participate in the lottery

[ ]

Thank you very much for participating!

The winner of the prize will be contacted in August.
For any questions, please contact me on: frerooy@hotmail.com

9.5.2 Control group

Dear responder,

Thank you for participating!

I am currently investigating people's eating behaviour and the dietary choices they make. This survey will not take you more than **10 minutes** to answer. One respondent will be randomly chosen and win **€20**.

I would like to notify you that your answers are anonymous and will be treated carefully. Your data will stored only for the purpose of this research. If you wish to get the chance of winning the prize, you can fill in your email address at the end of the questionnaire in order to contact you in the future. Your email address will be used solely for contacting you about the prize.
However, if you wish to remain **completely anonymous**, it is also possible. In that case, you will not be able to win the prize.

First, you will get 10 questions about your current food and drink intakes.
Please choose the answer you identify most with and answer these questions **truthfully**.

Let the survey begin!

Do you eat breakfast every day?

- ○ Yes
- ○ No

Do you use sugar in your coffee or tea?

- ○ Yes
- ○ No

Do you eat whole grain bread, rice and pasta instead of white bread, rice and pasta?

- ○ Yes
- ○ No

Do you consume soft drinks (e.g. coca cola, Fanta, energy drinks) every day?

- ○ Yes
- ○ No

Do you eat instant meals at least once a week?

- ○ Yes
- ○ No

Do you eat at least two pieces of fruit per day?

- ○ Yes
- ○ No

Do you eat fried foods (e.g. chips, fries, croquettes) at least once a week?

- ○ Yes
- ○ No

Do you eat at least 200 grams of vegetables per day?

- ○ Yes
- ○ No

Do you order take-out at least once a week?

- ○ Yes
- ○ No

Do you eat at least one less healthy snack (e.g. cookies, chocolate, chips) per day?

- ○ Yes
- ○ No

You are halfway there!

Imagine that you are doing groceries for the next day and you have to determine what you will eat.
For every meal you have the choice between two alternatives. Both alternatives are equally priced, available in the same supermarket and have the same preparation time.

Let's choose!

For breakfast, I prefer:

(Soy) Yogurt with oatmeal and fresh fruit

Pancakes with maple syrup and fresh fruit

For lunch, I prefer:

Quinoa salad with roasted vegetables

Caesar salad with grilled chicken

As an afternoon snack, I prefer:

A handful of potato chips

A handful of pistachio nuts

For dinner, I prefer:

Cauliflower steak with sweet potato fries

Steak with French fries

As dessert, I prefer:

Strawberry cheesecake

Fresh strawberries

You are almost done! To finish the survey, please fill out the demographic questions below.

Are you vegetarian/ vegan?

- ○ Yes
- ○ No

What is your age?

[ ]

What is your gender?

- ○ Male
- ○ Female

What is your nationality?

[▾]

What is the highest level of education you have completed?

- ○ High school
- ○ College
- ○ Bachelor graduate (university)
- ○ Master graduate (university)
- ○ Doctorates degree
- ○ [ ] Other, namely

Please enter your email address if you wish to participate in the lottery

[ ]

Thank you very much for participating!

The winner of the prize will be contacted in August.
For any questions, please contact me on: frerooy@hotmail.com