ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis

Master of Science in Economics & Business

Major in Behavioural Economics

Using the BTS to find the Truth behind the Untruth

Author: Amielle Guilloux

Student number: 409431

Supervisor: Sophie van der Zee

Second reader: Aurelien Baillon

Study program: Business Economics

Specialization: Behavioural Economics

Date of final version: 29/08/2018

Abstract

In the lifetime of a person, several lies have been told to him/her or by him/her. However, the reason why these lies have been told varies significantly across situations. Through a self-report study, this paper explores the motivations behind people's altruistic lies and whether more selfish intentions are at the source of these lies. In addition, it attempts to see whether people lie about the reasons why they tell these altruistic lies. This was done through the use of an economic truth serum, known as the Bayesian Truth Serum (BTS), which induces people to tell the truth by financially rewarding them when they answer a survey truthfully. The survey contained two main questions: one which focused on the main motivation people had behind their altruistic lies and another which asked whether people had lied more than once in the past 24 hours. The 400 respondents gathered were divided into two conditions: one control condition and one subjected to the BTS. The goal of this research was to see whether being subjected to the BTS provided significantly different responses to the two main questions, compared to the control condition. In other words whether respondents not facing the BTS were dishonest. The results show that people not facing the BTS do not have significantly different responses than those facing the BTS. Therefore, no evidence is found indicating that people not facing the BTS were lying. Nevertheless, the BTS validates that in general people have more selfish motivations when telling altruistic lies and that the majority of people lie at most once in a day.

1. Table of Content

2. Introduction

As human beings, we are constantly trying to improve ourselves and strive to portray a positive image of our persona to the outside world. Engaging in pro-social behaviour, more specifically altruism, is one of the many different ways in which we attempt to do this. Altruism is identified as a selfless act, solely trying to increase the welfare of another person (Batson, 1991). However, the extent of this selflessness is often under debate. Thomas Hobbes (1651) believed *"no man giveth but with intention of good to himself"*. In other words, he did not believe in selfless acts and claimed that all altruistic deeds hide more selfish intentions. Considering, that pro-social behaviour is a way to improve oneself is paradoxically no longer a selfless act, but more as one attempting to serve oneself.

Barasch, Levine, Berman, and Small (2014) debated on the theme of altruism and whether emotional benefits that may come into play when an altruistic deed is done can penalize the act as no longer being altruistic but selfish. Altruistic justifications have also surfaced as a way to excuse borderline immoral acts (Shalvi, Gino, Barkan, and Ayal, 2015). That is, immoral acts can be deemed acceptable by virtue of altruistic motives, for instance, altruistic lies. This mechanism, which people use to reduce their internal moral conflicts with justifications was brought in through cognitive dissonance theory (Festinger, 1957).

In general, a lie possesses a negative connotation but according to the utilitarian perspective, which qualifies a lie solely based on the reason behind a lie, an altruistic lie should not be condemned with such a negative connotation (Bentham, 1843; Mill, 1869). What is seen in everyday interactions is that people tend to excuse their lies by means of pro-social motives (Kant, 1949). In line with ideology presented by Hobbes (1651) do altruistic lies, which sole aim are to improve the welfare of the other person, really exist? Are there not more selfish reasons driving the communication of such lies?

The issue with obtaining answers to these questions is that they involve certain types of controversial behaviour (i.e., behaviour that people may have difficulty in revealing), and therefore cannot be asked directly to a person without doubting the truthfulness of the answer provided. The rationale for this struggle people face in admitting the truth to questionable behaviour is due to social desirability bias (Fisher, 1993). Social desirability bias claims that people have the desire to present themselves in a favourable manner, in accordance to social norms. Therefore, due to this social bias, people may feel compelled to lie about their questionable behaviours. If people were asked directly if they had more selfish reasons driving their altruistic lies, they would most likely hide the real truth out of shame and also perhaps to appear favourably. Thus, to obtain truthful answers to these questions a lie detector should be implemented.

Several researches have been implemented on lies[1], such as the amount of lies people tell (DePaulo, Kashy, Kirkendol, Wyer, and Epstein, 1996; Serota, Levine and Boster 2010; Debey, Schryver, Logan, Suchotzki and Verschuere 2015; Feldman, Forrest, and Happ 2002) or even the type of lies that are told (Tenbrunsel & Messick, 2004; Bryant, 2008; Levine, Ali, Dean, Abdulla, and Garcia-Ruano, 2016; Morris, 2017). As will be further developed in the literature review, depending on the various truth-elicitation methods implemented (diary, self-report or experimental) in these papers, different results were obtained. Although, daily diary and self-report surveys are simpler and quicker to conduct compared to experimental research, nevertheless they have a risk of lacking in accuracy due to social desirability bias (Fisher & Katz, 2000). This paper will attempt to improve the accuracy of self-report surveys with the inclusion of economic lie detecting device, known as *The Bayesian Truth Serum* (BTS).

Drazen Prelec (2004) introduced an economic truth serum, known as *The Bayesian Truth Serum* (BTS). He presented a scoring method, which induces people to tell the

[1] These papers will be discussed in the Section 3

truth by financially rewarding them when they answer a survey truthfully[2]. The motivation of this paper is to explore whether the BTS can be utilized to uncover the truth about people's untruthfulness. Inherently, it will attempt to observe whether the BTS is an effective tool in obtaining the truth when asking questions directly related to lying attitudes. If this is confirmed, it will allow for other researches focusing on lying behaviours to use the BTS as a lie detector. Due to this new practical utilization of the BTS, the scientific relevance of this study is advocated and could therefore provide further research opportunities in this domain.

So far the BTS has mainly been used in treating issues, which people may be uncomfortable in revealing due to social desirability bias, such as cheating, stealing or even racism (Abolmagd, 2017; Schneider, 2017). These previous applications of the BTS are strongly linked to lying and show promising results in using the BTS to solely look into peoples' lying behaviours. This will allow a better comprehension behind the psychological considerations behind people's lying behaviours and what moral circumstances influence people to lie about the motives of their lies. The latter provides the social relevance of this research.

The current study will isolate lies allegedly told for the benefit of others (i.e. 'altruistic' lies). Firstly, it will examine if altruistic lies are truly legitimate. That is, are people truly selfless when telling altruistic lies or are there more selfish motivations behind these types of lies. It is expected that there will be a divergence between the reason people claim to altruistically lie to others and the actual motivation behind their lies. As DePaulo et al. (1996) pointed out: "lies are more often told to serve the self than to benefit others". As DePaulo et al. (1996) pointed out: "lies are more often told to serve the self than to benefit others". Secondly, it will attempt to unravel whether people lie about these motivations. In other words, are people ashamed in admitting what their true motives are for telling such lies and therefore are compelled to lie about them? The experimental set up will consist of a control group (i.e., people undergoing the normal version of survey) and a treatment group

[2] In depth explanation of the BTS is provided in Section 3.5

(i.e., people undergoing the BTS version of the survey). The results of both conditions will then be compared in order to see if respondents in the treatment condition provided significantly different answers to respondents in the control condition.

This thesis is divided into seven sections: (2) Introduction, (3) Literature Review, (4) Methodology, (5) Data, (6) Results, (7) Discussion & Limitations and (8) Conclusion. The Literature Review goes through an overview of various lie definitions and classifications. It will also examine the lie frequency and discuss lie detection techniques as well as present the BTS. The Methodology consists of five subsections, which will explain the experimental design, the BTS algorithm, the survey design as well as the participants targeted and finally the statistical analysis, which will be applied. The Results section will go through each statistical test implemented. The Discussion and Limitations section will comment on the relevant output obtained, the potential direction of future research, and discuss the main limitations of the study. Finally, the Conclusion section will end this paper presenting the central findings of the study.

3. Literature Review

3.1 Definition of a Lie

Deception or lying has been defined in various ways. Serota et al. (2010) describe it as the act of seeking knowingly and intentionally to mislead others. Later on, Serota (2015) defined lying as means to an end, in other words it serves as tool to facilitate access to our goals. Bergstrom (2009) characterized lying as a form of manipulation. He claimed that communication is a channel for deception and if a person is able to communicate that person is also able to manipulate. According to these diverse conceptions, a lie is *communicating* by not telling the truth. However, as Aldert de Vrij (2000) cleverly pointed out a lie can occur without the use of words. Lying by omission is a common example where the deceiver is lying by not communicating. De Vrij's (2000) formal definition of a lie is "*a successful or unsuccessful deliberate*

attempt, without forewarning, to create in another a belief, which the communicator considers to be untrue."

Nevertheless, there seems to be a digression between what is defined as lie and what people qualify as a lie (Fu, Cameron, Heyman, & Lee, 2007). For long periods of time, philosophers have debated this issue. On one hand, supporting de Vrij's (2000) definition, St. Augustine (1952), Kant (1949), and Bok (1999) considered that the act of lying, regardless of its purpose, is one that is undertaken consciously by the speaker whose intention is to make a false statement. That is the deceiver must know that the information he is imparting is false. According to the above-mentioned philosophers a false statement claimed by a person who believed the statement to be true would not be considered a lie. For example, a person suffering from a delusional disorder and asserting a false-hood they believed to be true would not be categorized as a lie.

On the other hand, utilitarian perspective, represented by philosophers such as Bentham (1843) and Mill (1869), claims that whether or not a false statement is consciously told, what qualifies it as a lie depends on the motive. If the deceit being told to a person is to increase their happiness or spare them pain then this should not be labeled as a lie. This view defines a deceit through its moral and social implications, consequently lying for the benefit others would not be condemned with such a negative connotation. As will be shown in the sub-section 3.3, discussing lie frequency, people do not have an accurate estimation of the amount of lies they tell. One possible explanation for this phenomenon may be caused by people no longer registering their social lies negatively (Backbier, Hoogstraten, and Terwogt-Kouwenhoven, 1997).

All of the above-mentioned definitions offer a brief insight that people may have various interpretations of a lie. Consequently, a precise definition will be offered in the introduction of the survey in order for the responders to hold the same definition in their minds. This paper will follow the definition offered by de Vrij (2000), as he seems to have the most complete and accurate perspective of what

constitutes a lie. De Vrij (2000) also included non-verbal lies in his definition, however since the main focus will be on communicated lies told for the benefit of others (also know as altruistic deceits) non-verbal lies will be disregarded.

3.2 Lie classification

A lie can be thought of as a conception that not only possesses various interpretations but can also be categorized into different classes. The different lie classifications allow a better understanding of the motivations behind lies.

One way to classify a lie is by looking at the reason why it is told. There are several reasons why people lie: to embellish their image, to hide bad behaviour or even to avoid conflict. A lie is a complex concept that can be employed for various goals. Morris (2017) identified four major categories based on a previous research undertaken by Levine et al. (2016):

1. To promote yourself
2. To protect yourself
3. To impact others
4. Unclear

Levine et al. (2016) conducted a self-report survey asking open-ended questions to the respondents. The purpose of the research was to develop a list of deception categories common across cultures by asking subjects to describe an instance of deception from both the perspective of the liar or the one being lied to. From the information gathered from Levin et al.'s (2016) investigation, Morris (2017) obtained that 'lying to promote yourself' explained the majority of why people lie. An example would be lying in order gain financial benefits or embellish your image. 'Lying to protect yourself' constituted the second most common reason of why people lie. Covering up a bad behaviour or a misdeed would be considered as such a lie. 'Lying to impact others', are the third most common reason why people lie. This can be a lie to avoid hurting a person's feelings or on the contrary to hurt a person. Since the current study isolates lies allegedly told for the benefit of others (i.e., altruistic lies),

it will therefore solely contemplate lies classified in this category. As their description suggests, lies 'benefiting others' are told in attempt 'to impact others'. Finally, the 'Unclear' category constituted least frequent reasons as to why people lie. Lies whose motives are unclear or are pathological (e.g. compulsive) fell under that category.

From another perspective, Bryant (2008) unlike Morris (2017), who constructed the separate categories through the results he obtained, Bryant (2008) conducted a study to view how people classified different types of lies. His results showed that depending on the intention, consequence, beneficiary, truthfulness and acceptability, a lie could be classified into three distinct categories:

(1) Real Lies

(2) White Lies

(3) Gray Lies

In this study, *'Real Lies'* were defined as being malicious, deliberate, deceptive, deceitful and serving purely egoistical needs. In Bryant's (2008) study they were classified as being unacceptable. On the other hand, *'White lies'* were viewed as being harmless, trivial and used in the majority of cases for altruistic reasons (e.g. protecting another person from harmful truth). Consequently, Bryant (2008) classified them as being socially acceptable and justified. Finally, *'Gray Lies'* were perceived as standing at mid-point between *'Real Lies'* and *'White Lies'*. Although DePaulo et al.'s (1996) results also found there to be three types of lies, nevertheless the labels for each type of lie differed. The three types of lies were labeled as: outright lies, subtle lies, and exaggerations. DePaulo et al. (1996) defined 'outright lies' as complete falsehoods "in which the information conveyed is completely different from, or contradictory to, the truth". This category comes closest to the 'real lies' identified by Byrant (2008), which seem to have the most negative undertone. 'Subtle lies' were referred to as "those told by evading or omitting relevant details and by telling literal truths that are designed to mislead" (DePaulo et al., 1996). Finally, 'exaggerations' are the ones where the liar simply overstates or extends the truth of a statement. Bryant's (2008) categories seem to be

distinguished through moral and social implications. For instance, 'real lies' have a strong negative undertone compared to 'white lies', which seem more socially acceptable and 'innocent'. However, DePaulo et al.'s (1996) classifications are not distinctive to one another in their moral or social implications. All the three lie categories can all have a positive as well as a negative implication depending on the motive behind the lie.

When focusing on lies that directly impact others, these too can hold both a positive and negative undertone. They can be told for purely selfish reasons with large negative consequences or on the contrary told purely for altruistic purposes to benefit others. The latter then justifies the act of lying in such a context, as being socially acceptable and hence to a certain extent 'morally' acceptable. But is this really the case? When we claim to lie for the benefit of others (i.e., for altruistic reasons), are our claims truthful? Could we be justifying our 'immoral' act through a 'moral' motive when in reality the true reason behind our lie is to satisfy egoistical needs? This echoes Shalvi et al.'s (2015) pre- and post-violation moral justification schematic (See Appendix A.2). In their paper they point out how a person's "moral self-concept is threatened at two points in time: before committing a moral violation (when ethical dissonance is anticipated) and afterward (when ethical dissonance is experienced)" (Shalvi et al., 2015). Moreover, they explain how self-serving justifications can be used to decrease a person's guilt and avoid their moral self-concept to be tainted. Thus, we are then able to protect the image that we have of ourselves from any form of value deterioration that such immoral acts could have. Mazar, Amir, and Ariely (2008) named this the fudge factor, which is simply the ability to misbehave and still think of ourselves as 'good' people. The current paper will also, like Shalvi et al. (2015) look into immoral acts but more precisely, it will focus on the immoral act of lying and look into whether altruistic lies are truly provoked by selfless intentions.

3.3 Lie Frequency

If you were to ask the average person if they considered themselves an honest person, they would most likely agree with the statement (Ariely & Melamede, 2015). However, research shows that on average we lie more often than we think. Indeed, DePaulo and al. (1996) in their daily diary reports found that the average person lies on average between one to two times in a day.

Through a self-report study, Serota et al. (2010) and Debey et al. (2015) re-investigated the claim made by De Paulo et al. (1996) and although they obtained the same average results (i.e., that on average people tell one to two lies in a day). Nevertheless, they found the frequency of the lies to be non-normally distributed around the mean. Serota et al.'s (2010) results (see Appendix A.1) established that 50% of all lies told in their study, were only told by 5.30% of the respondents. In parallel Debey et al. (2015) obtained that out of the 50.67% of all the lies reported, were told by a minority of 8.87%. These minorities, in both studies, were labeled as prolific liars.

In an experimental research setting attempting to identify how much people lie, Feldman, Forrest, & Happ (2002), studied how self-presentation goals impact the amount as well as the type of lies we tell when engaged in a 10-minute conversation. Feldman et al. (2002) retrieved that 60% of their respondents lied within the 10 minutes with an average of 3 lies, instead of one to two times in a day. Contrary to DePaulo's (1996), Serota et al.'s (2015) and Debey et al.'s (2015), Feldman et al.'s study treated self-generated lies that were identified just after the deception was undertaken. Therefore, the lie frequency prediction was most likely more accurate in the experimental research than in the self-report and diary studies. However, Feldman et al. (2002) noted that a possible cause for this unexpected amount was perhaps due to forcing participants to engage in a conversation with a complete stranger, which may have driven them to exaggerate or even invent stories in order to deal with the awkward position they were put through. One occurring that may invalidate this critique is that we have to interact with strangers on a daily basis. Nonetheless, although the latter phenomenon is true, most of the times we choose

to interact with these strangers and consequently have already established a subject of conversation prior to engaging with the stranger in question.

The one thing all of the mentioned papers had in common is that their results relied on people's honest self-evaluation. That is the researcher had to trust that respondents were truthful in the responses they provided on their lying behaviours. Therefore, this makes them accountable to suffer from under-reporting or social desirability Bias (Fisher & Katz, 2000). In the majority of cases, when people are asked to report deceptive or questionable behaviour, the usual tendency is for them to underestimate such behaviour due to social norms, which is known as social desirability bias. Biased answers are common in self-reports for social desirability or social approval (Arnold & Feldman, 1981). In the current self-report study, the BTS will be used to reduce social desirability bias since, as will be explained later on, it is in the respondents best interest to tell the truth as it constitutes the Bayesian Nash Equilibrium. Furthermore, since there will be two conditions: one BTS condition and one control condition, if respondents under-report in the control condition it will be apparent when comparing their answers with the answers from the BTS condition. If the BTS proves to reduce social desirability bias, then this would allow for a new way to implement self-report studies, allowing for more reliable results to be obtained. Increasing the reliability of self-report studies would be extremely useful as they are cheaper and easier to undertake than experimental researches. In addition to investigating the motivations behind lies, this study will also incorporate a question regarding lie frequency, in order to see whether or not self-report surveys do indeed suffer from Social Desirability Bias. A direct comparison of the responses obtained in this experiment with the ones in the literature will not be possible since the questions in the latter were open ended and the questions in the BTS need to be multiple choice. However, it will be possible to see if the amount of people admitting to have lied more than once in a day, in the previous studies, corresponds to the same about obtained in this study.

All of the previous studies, in their discussions, highlighted that people lie more than they think and have a tendency to underreport their lying behaviour. We have an

inclination to form an ideal scenario in our heads and frame our beliefs around this ideal scenario (Ariely & Melamede, 2015). To reiterate, we may frame our beliefs of our lie frequency based on an ideal conception we have. Similarly, in line with the reasoning by Bentham (1843) and Mill (1869) in our current society, morality may still depict whether a statement is considered a lie causing some lies not to be labeled as 'lies'. For instance, White lies no longer seem to have a moral impact on the majority of the population, "because they are trivial and may even prevent someone from being hurt by an unnecessary truth" (Bryant, 2008). In other words, a moral adaptation has occurred allowing these lies to be 'socially' acceptable and no longer be viewed with such a negative connotation (Backbier et al., 1997). Therefore, when telling white lies people experience little or no feeling of discomfort and if asked if they would replicate this lie in another scenario, 70% would (DePaulo, 1996) indicating that a moral adaptation may have occurred. The feeling of discomfort can be observed through the activation of regions in the brain involved with emotions such as the Amygdala and the Insula. If you qualify a type of lie as being 'morally bad' there will be a high response in such regions of the brain. Hence, when white lies are told, the emotive regions of your brain will not respond, as the negative feeling (i.e., the guilt) no longer prevails (Ariely & Melamede, 2015). Due this moral adaptation, you will not register or classify it as a lie. Consequently, on average and without much thought people will report themselves as being honest or even may not consciously be aware that they are lying. Therefore, underreporting may occur not only due to social desirability bias, but also because people don't mentally register their lies and not even realize that they are under-reporting their lie frequency. Although, this could perhaps also be an issue in the current research, even with the use of the BTS, nevertheless it may be avoided by asking about the behaviour others may display, which encourages people to carefully analyze themselves and more importantly focus when answering the questions.

3.4 Lie Detection

Although, using deception seems to be innate to most human beings, people seem extremely poor at detecting it (Bond & DePaulo, 2006). For centuries several

methods have been invented attempting to detect dishonest behaviour. In 1000 B.C., the Chinese implemented a method where they asked their suspects to place rice in their mouths and after a couple of seconds spit it out. If the rice remained dry then the person was condemned as a liar. This belief came from a physiological phenomenon when in face of fear or anxiety, consisting of a decreased salivation, which leads to a dry mouth (Ford, 2005). Later this link between biological perturbations and deception came to be more explicitly exploited, for instance by the Polygraph. Indeed, physicians such as Lewis Thomas (1983) asserted: "a human being cannot tell a lie, even a small one, without setting off some kind of smoke alarm". The idea is that a falsehood causes unease within beings to which the body responds and ultimately sending signals that can be detected. This constituted the foundation upon which current lie detectors were created.

Due to the increasing interest in detecting deceits different approaches have been undertaken for that purpose, including blood pressure tests, psychological stress evaluations, truth serums or even hypnosis. These lie detecting mechanisms become extremely useful in several social as well as economic situations, such as criminal interrogations, interviews or even surveys. The reliability of these practices however, are constantly being questioned (Burkey, 1965), which has served to encourage researchers to improve or come up with new ways to obtain the truth.

3.5 The Bayesian Truth Serum

The BTS is an economic truth serum introduced by Drazen Prelec (2004). The main goal of this truth serum is to motivate respondents to tell the truth by rewarding them for their honesty. An algorithm assigns respondents a score based on their degree of truthfulness for the answers provided, as well as their predictions of other respondents' answers.

When the BTS algorithm was first presented its main purpose of application was to elicit trustworthy information from surveys in "studies of public or expert opinions: voting intentions, product rating, expert forecasting and various crowdsourcing

applications" (Cvitanić, Prelec, Radas, & Šikić, n.d). So far, the BTS has been implemented in several different experiments, for instance assessing the reliability of deterrence theory (Loughran, Paternoster, & Thomas, 2014), evaluating the knowledge gained in design courses (Miller, Bailey, & Kirlik, 2014), improving the choice based conjoint (Dimitrov, 2017) or even incentivizing digital pirates' confessions (Kukla-Gryz, Tyrowicz, Krawczyk, & Siwiński, 2015).

Although some BTS papers (Prelec, 2004; Cvitanić et al., n.d.; Baillon, 2017) have centered their research on the algorithm in order to validate its efficacy, the papers (Loughran et al., 2014; Dimitrov, 2017; Kukla-Gryz et al., 2015) actually making use of the BTS as a truth serum have only used the algorithm as a way to reward their participants, since it would allow them to see who provided the most truthful answers. Therefore, other than rewarding participants, the algorithm is not the predominant function of the BTS. The main way in which the BTS is applied in studies is as an experiment condition to compare with other conditions and consequently see if a difference between the responses in each condition is significant. This will also be the approach implemented in this paper in order to see if people in the control condition display dishonest behaviour.

3.5.1 Mechanism and intuition

In a BTS survey, responders are faced with a multiple choice question, where they not only have to provide their personal answer but also the percentage estimate of how other responders will respond. The basic mechanism of the BTS scoring method is that it compares the respondent's personal answer with their prediction of what others would respond. Through this, it can determine how truthful a person was and rewards the respondent accordingly. The BTS score of a person is calculated based on two factors: (1*) the information score* and *(2) the prediction score*.

(1) The Information Score

The first part of the BTS, the *information score* is calculated based on the respondent's answer and the probabilistic prediction of what other respondents will answer. For instance, the second question in the survey will be: *"Have you lied more than once in the past 24 hours?"* to either a *'yes'* or *'no'* answer would be provided. The question following the first question would then ask: "*What is the probability that the other respondents would have lied in the past week?*" Here a percentage estimate would be answered.

The information score formula for answer k is the following (Prelec, 2004):

$$Information\ score = \ log \frac{\bar{x}_k}{\bar{y}_k}$$

The information score for respondent r for each k number of answers:

$$Information\ score = \sum_k x_k^r \ log \frac{\bar{x}_k}{\bar{y}_k}$$

Where $\bar{x}_k$ is the actual average frequency for answer k and $\bar{y}_k$ is the average of the predicted frequencies for answer k. Note that x_k^r will be $= 0$ or all answers apart from the one chosen by the responder, where it will equal $= 1$. An answer receives a high score, when the actual average frequency of that answer is greater than its average predicted frequency. These types of answers are labeled as 'surprisingly common' answers. Answers where the opposite holds (i.e., the predicted frequency is greater than the actual frequency) are labeled 'surprisingly uncommon' and receive low scores. For example, consider a binary question (e.g. Yes/No). If the actual mean frequency of the 'Yes' answers is 20% and the collectively predicted frequency of the 'Yes' answer is 15% then the respondents who provided the answer 'Yes' obtain a high score, seeing as this is a 'surprisingly common' answer. However, if the collectively predicted frequency were 25% and the actual frequency remained 20%, people answering 'Yes' would now obtain low scores. To reiterate, when the ratio is > 0 then the information score will be positive but if the ratio is < 0, then the information score is negative, thus penalizing the respondent. Scores obtained

are an indication of a respondent's degree of honesty. Since the BTS is always accompanied with an incentive scheme that rewards the most honest answers, if the respondent wishes to exploit prize as much as possible they will need to provide truthful answers in order not to be penalized by low scores.

The main intuition behind the common and uncommon answers is that a person will tend to bias their belief of what other people's answers are for a question, based on what they answered for the same given question. Ross, Greene and House (1977) mention this phenomenon as a 'false consensus' or 'egocentric bias'. In other words, people perceive their choices as being relatively common and any deviation from that default is considered uncommon. This can be linked to the availability heuristic explored by Kahneman and Tversky (1973) by which people evaluate frequencies or probabilities of events occurring by the availability of information they have. Consequently, when giving your probabilistic estimate of what other respondents would answer you are framing your prediction based on your true opinion/answer. This is because your true opinion/answer is the most available piece of information that comes to mind. Prelec (2004) named this the common prior and claimed it to be a necessary assumption in order for the truth serum to work. Providing an estimate of other people's behaviour leads to a more reliable insight into people's true opinions since a distancing factor is established by not evaluating oneself but a separate party (Fisher, 1993).

It is essential to note that, although the "BTS exploits the subjective correlation between a person's opinion and the opinion of others" (Weaver & Prelec, 2013), truthful answers in all circumstances will grant respondents with high scores making it in their best interest to answer truthfully, since respondents' scores are tied to their (financial) rewards. That is, even if the respondent has an uncommon answer relative to the rest of the sample, they will not be penalized. BTS is the reverse of *consensus scoring* which favours the most popular answer influencing people to mold their answer with respect to what they expect others will answer and hence creating incentives to deceive (Weaver & Prelec, 2013). Therefore, truthful reporting

in the BTS constitutes the Bayesian Nash equilibrium as it provides the highest expected value (Baillon, 2017).

(2) The Prediction Score

The second part of the BTS, the prediction score can be thought of as the relative entropy or the Kullaback-Liebler divergence (1954), which is a correction factor to any divergence between two probability distributions. In other words, the prediction score corrects for any deviation between the actual average frequency ($\bar{x}_k$) of an answer and a person r's prediction of that distribution(y_k^r). It is an element, which penalizes the deviation the prediction (y_k^r) has from the actual distribution. The best prediction score a person can have is zero, i.e., when $y_k^r = \bar{x}_k$ (Prelec, 2004). That is, if a person has a prediction score of zero for a question, then that person estimated the exact percentage amount, of the answer to that question was selected by the other respondents. When the two probabilities diverge a negative prediction score results, which penalizes the overall BTS score. The prediction score can be thought of as a score assigning respondents a grade for how well they are able to predict what the actual behaviour of the other respondents is.

Prelec (2004) demonstrates the formula as follows:

$$Prediction\ score = \ \alpha\ log \sum_k \bar{x}_k log \frac{y_k^r}{\bar{x}_k}$$

Where α is a constant that "fine-tunes the weight given to the prediction error" (Prelec, 2004).

(3) BTS algorithm

Combining the information score and the prediction score, we obtain the overall BTS score:

$$BTS\ score = \sum_k x_k^r\ log \frac{\bar{x}_k}{\bar{y}_k} + \ \alpha\ log \sum_k \bar{x}_k log \frac{y_k^r}{\bar{x}_k}$$

The responses regarded as most 'truthful' obtain a higher score and this is viewed when the actual frequency is higher than the predicted frequency. For the BTS to be successful, respondents do not need to understand how the algorithm functions, just that the truth rewards them with high scores, which in turn increases their chances of receiving a financial or non-financial reward. The BTS algorithm has been thoroughly investigated and validated by several researchers (Miller et al. 2014; Weaver & Prelec 2013; John, Loewenstein, & Prelec, 2012; Loughran et al., 2014).

4. Methodology

4.1 Experimental and Survey Design

To investigate intro people's lying behaviours an experiment was undertaken. More precisely, respondents of the experiment were asked to fill in an online survey, conducted through the use of the Qualtrics platform. The survey was distributed via social media, namely Facebook and Instagram, using an anonymous link. Participants were also approached in real life on university campus. Since the questionnaire was created using the Qualtrics software, it was necessary for participants to have access to an online device (e.g. smartphone, computer or tablet) whilst completing it. Furthermore, ethical approval for this study was granted, which required respondents to be least eighteen years of age.

The survey was constructed as a between-subject design with two versions of the survey: BTS and Control version. Participants who completed the BTS version of the survey constituted the treatment group and those who completed the control version of the survey served as the control group. Each participant was allocated randomly to respond to either version through an automated feature available on the Qualtrics platform. The probability distribution of the conditions was 50%, meaning that people taking the survey would be equally distributed across the conditions.

The function of the BTS in this study was to see whether a divergence in the answers obtained between the two conditions would be apparent. More specifically, whether people assigned to the control condition would offer answers, which are considered more socially acceptable, than people assigned to the BTS condition. Consequently, the condition (BTS or Control) would constitute the independent variable and the answer provided to the two core questions (i.e., *"What was the main why you lied?"* & *"Have you lied more than once in the past 24 hours?"*) would establish the dependent variable.

4.2 Experimental Procedure

In both conditions, the respondents were first faced with a preliminary introduction of the experiment (See Appendix B.1.1). This introduction informed them that by participating in this survey they had the chance of winning €20, but in order to engage, they needed to provide their email address at the end of the survey (See Appendix B.1.3.). In the BTS condition the winner of the BTS was determined based on the BTS score. That is, participants were made aware that their responses would be evaluated using a scoring method which assigns higher scores to truthful answers and that the participant obtaining the highest score would win the €20 (See Appendix B.1.4.). However, in the control condition the winner of the €20 was chosen by means of a random lottery. In both conditions it was stressed that contact information would solely be used for the prize delivery and nothing else and that if they wished to remain completely anonymous this was also possible but in that case they would not be eligible to win the €20. In addition, it was also mentioned that the email addresses would be deleted once the prize was delivered.

When undertaking a BTS experiment the dilemma of anonymity is always faced. On the one hand, rewarding people for providing truthful answers is essential. Nevertheless, when questionable behaviour is the center of the research or when the experiment considers information, which people may have a hard time admitting, anonymity is important (Abolmagd, 2017; Dimitrov, 2017; Schneider,

2017). When using the BTS, in order to reward a participant for their truthfulness, their answers need to be tied to their contact details. This may result in people being reluctant to actually answer honestly due to the sensitive information that may be revealed. To overcome the anonymity dilemma, several papers have undertaken the charity donation approach (Abolmagd, 2017; Dimitrov, 2017; Schneider, 2017). Instead of rewarding the participant directly for their truthfulness, the responder had possibility to choose one charity project, which will receive the amount they have gained. However, this approach was not implemented in this experiment. The reason for this, is that people may not feel truly incentivized to provide truthful answers, as the reward is not bound directly to them. In other words, respondents may not feel directly impacted by providing truthful answers. Therefore, anonymity was chosen for this study, as it was essential to maximize the chances of obtaining truthful answers. Hubbard and Little (1988) support this claim as they found the efficacy of charitable donations to be weak. Likewise, Warriner, Goyder, Gjertsen, Hohner, & McSpurren (1996), building from Hubbard and Little (1988) incentive review, found identical results. Charitable donations would not be an effective tool in incentivizing people to provide truthful answers constitute, consequently, another approach was used. In this experiment, people had the opportunity of supplying their email addresses if they wished to have the chance of winning €20 and could also remain fully anonymous if they felt uncomfortable about the responses they had just given. The email provision box appeared at the end of the survey allowing for participants to be aware of the sensitive questions asked to them prior to giving their contact information (See Appendix B.1.7.). That is the decision of *remaining* anonymous would only be done at the end of the survey once all answers were provided. Therefore, throughout the survey, participants would answer as if they were anonymous and only at the end of the survey they could choose if they were comfortable with supplying their email address or not.

Prior to the main questions, four demographic questions were asked (age, gender, nationality, and relationship status) in order to get a general impression of the sample composition (See Appendix B.1.2). Next, participant were provided with a brief introduction and explanation of people's lying behaviour as well as three

examples of what was considered as an 'altruistic' lie. This explanation was identical in both versions and was provided to avoid any hesitations the respondents may have regarding what a lie is and what constitutes an altruistic lie, in this experiment (See Appendix B.1.5-B.1.6). In the BTS condition, there was a supplementary part of instructions explaining to the respondents that a scoring method would be used to evaluate their truthfulness and that telling the truth would maximize their score (See Appendix B.1.4.).

Thereafter, participants were asked in both conditions, whether they had ever lied to spare someone's feelings, as this was a requirement needed to continue the rest of the survey. Consequently, the survey would end for respondents who had answered 'No' to that question. The Qualtrics platform offers the option to add conditions depending on which answer was provided. After answering if they had ever lied to spare someone's feelings, participants were presented with an additional explanation, highlighting how an altruistic lie can have a more subtle and egoistical motivation, followed (See Appendix B.1.5- B.1.6). The three examples were chosen deliberately, as although they are usually claimed to be altruistic deceits they could just as well hide a more selfish intention. For instance, *"Telling a person 'those pants look great' when they don't"* is a type of lie that may be told to preserve the feelings of the one being lied to. Conversely, the true underlying reason may be to refrain from conflict, which would put the liar in an awkward position or even perhaps cause the liar to be disliked by the other person constituting more selfish intentions.

Ensuing the instructions, participants were faced with the two core questions of the experiment. The first core question was a follow up on the question that had been asked just before (i.e., *"Have you ever lied to spare someone's feelings?"*). It asked, respondents to answer honestly what was the ***main*** reason why they lied (i.e., *"What was the **main** reason why you lied?"*). Two options were offered:

(1) Spare their feelings

(2) Spare yourself from being in an awkward position, conflict or the

eventual loss of the person.

The **'main'** was highlighted in red (see Appendix B.1.5-6) since usually people would argue that it would be a combination of both options. Additionally it was stipulated that they should refer to the lie situation that appeared in their minds when they answered the question *"Have you ever lied to spare someone's feelings?"*

As there is extensive literature concerning lie frequency (DePaulo, 1996; Serota et al., 2015; Debey et al., 2015), and these studies have indicated signs of suffering from social desirability bias, consequently, the second core question asked if respondents had lied more than once in the last 24 hours (i.e., "*Have you lied more than once in the past 24 hours*?). To this question a simple Yes/No option was offered. Moreover, the answers obtained to the main questions would be the base upon which the average *actual frequency* would be calculated ($\bar{x}_k$). A direct comparison with the responses obtained in this experiment with DePaulo's (1996), Serota et al.'s (2015) and Debey et al.'s average of two lies in day, will not be possible due to the way in which the BTS question was structured as a multiple-choice question, for calculation purposes. However, it will be apparent if the results from these previous self-report studies did in fact suffer from social desirability bias, if they diverge significantly from the ones obtained in this experiment.

In addition, in the BTS version the two core questions were followed by a question, which would ask to predict the percentage of people who would choose the same answer as them. These questions were answered through the use of a slider where respondents could choose the value between 0-100%. It is important to note that any response, which that had values 0 or 100 were changed to 1 or 99 since the value of 0 or 100 are incompatible, due to the logarithmic transformation, and thus inefficient with the BTS scoring method. The answers provided to these questions were then used to calculate the average *predicted frequency* ($\bar{y}_k$).

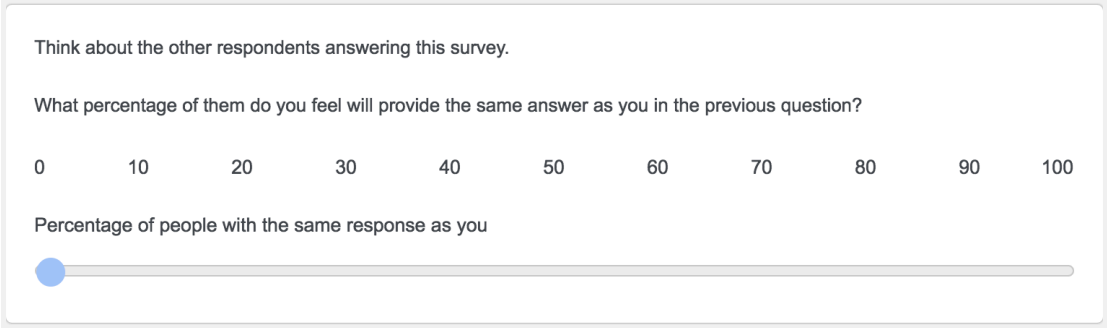


Figure 1: Slider Option

The answers obtained for both conditions will be compared to test if a significant difference prevails between them. More specifically, a Pearson-Chi square test will be undertaken to analyze whether the people in the control condition provide significantly more socially desirable answers compared to those who faced the BTS version of the survey. A significant difference between the answers in the two conditions would be an indication that the people who constituted the control condition were not entirely honest in their answers and that the BTS made people in the treatment condition more honest. This interpretation is done due to the extensive literature supporting the validity of the BTS and its effectiveness. Furthermore, in order to guarantee that respondents behave according to the Bayesian Nash Equilibrium (i.e., that respondents state the truth), the common prior assumption has to be met. If this holds, then this would signify that the truth-telling incentive mechanism functioned well and would imply that the BTS condition contains more honest than dishonest answers. Hence, the results of the BTS would be used as representation of the truth and any deviation from this reference would constitute a 'lie'. The application of the BTS could provide a new insight on how to reduce social desirability bias in self-report studies.

4.3 Statistical Analysis

As has been shown from the papers investigating and validating the BTS, its main function is that it promotes more truthful answers compared to a normal survey (Prelec, 2004; Weaver et al., 2013; Abolmagd, 2017; Baillon, 2017; Cvitanić et al., n.d). Consequently, the data gathered from the BTS version will stamp the 'real'

common lying behaviour of the average population. The control condition will serve to test whether people normally tend to hide the real truth about the reasons why they lie. If a significant difference in the average responses is found between the two conditions then we can conclude that people are not being truthful concerning the questions that were asked to them. Such a conclusion is only possible, if the validity of the BTS is confirmed. In order for the truth serum to be valid, the estimated frequency of the people who admit for example to lying for selfish reasons, must be significantly different to the people who declare lying for more altruistic reasons. Higher percentage means for the former behaviour (i.e., lying for selfish reasons) are expected for the people who claim to lie for selfish reasons than for the people who claim to lie for altruistic reasons (Thomas Hobbes, 1651; Festinger, 1957; Shalvi et al., 2015). This constitutes the common prior assumption asserted by Prelec (2004). Hence, in the BTS condition, a t-test is applied for the percentage estimation questions to examine if the mean of the percentage estimate regarding the questionable behaviour (i.e., admitting to have lied for more selfish reasons or to have lied more than once in the past 24 hours) is significantly different between the people admitting to this behaviour than the people who do not. It is important to note that in the questionnaire the percentage estimation question asked *"What percentage of them do you feel will provide the same answer as you in the previous question?"* Therefore, if a person provided the answer *'spare their feelings'*, their percentage estimation was subtracted from 100 and the people who provided the other answer their percentage estimation remained the same. The same was implemented for the question, which asked whether a person had lied more than once in the past 24 hours (Q2). However, for Q2 the percentage estimation of the people who had answered *'No'* was subtracted from 100 and the percentage estimation of the people who had answered *'Yes'* was preserved.

The Pearson Chi-Square test will be implemented to compare the results in both conditions. It tests whether a significant difference in the response chosen occurs depending on which condition was assigned. In other words, does belonging to one the BTS condition drive participants to choose one answer compared to if they were in the control condition. To facilitate the running of the test, a dummy variable

Control was generated. This variable takes the value of 0 when referring to the control condition (i.e., Non-BTS version) and takes the value of 1 when referring to the treatment condition (i.e., BTS version).

Likewise, for both main questions (i.e., *"What was the main reason why you lied?"* & "*Have you lied more than once in the past 24 hours?"*), two dummy variables were generated. These take the value of 0 if the first option was selected and 1 if the second option was chose by a participant. For the variable Q1, the first option was the response "*Spare their feelings"* and the second option was "*Spare yourself from being in an awkward position, conflict or the eventual loss of the person".* For the variable Q2, the first option was the answer 'Yes' and the second option was the answer 'No'.

According to Thomas Hobbes (1651), Festinger (1957) and Shalvi et al. (2015) it is hypothesized that people in the BTS condition will admit to having more selfish reasons behind their altruistic deceits, this constitutes the first hypothesis (H1). Likewise, relying on what the literature has uncovered (DePaulo et al., 1996; Debey et al., 2015), it would be expected that people on average lie more than once in 24 hours (H2). Conjointly, the BTS will serve to test whether people have a tendency to lie about their lying behaviours. This belief that people may be inclined to lie about their lying behaviours stems from the concept of social desirability bias (Fisher, 1993). Social desirability bias demonstrates people being uncomfortable in admitting the truth about their questionable behaviours. It is expected that respondents in the control condition will provide more socially acceptable answers than the respondents in the BTS condition. Consequently, when asked whether more selfish motives are at the source of their altruistic lies, it is expected that people in the control condition will lie about it more than people in the BTS (H3). Likewise, in line with DePaulo et al. (1996), Serota et al. (2010) and Debey et al. (2015), it is predicted that on average, compared to the people in the BTS condition, people in the control condition will underreport their lie frequency by claiming they have not lied more than once in the past 24 hours. These hypotheses are proven, if a significant

difference is found between the mean frequencies of the BTS and control condition. In sum, the following hypotheses have been formulated:

H1: On average, people lie more than once in 24 hours

H2: On average, people's real motives behind altruistic lies are more selfish than selfless

H3: On average, participants in the BTS condition will report to have lied for selfish reasons, more times than participants in the Control condition.

H4: On average, participants in the BTS condition will report to have lied more than once in the past 24 hours, more times than participants in the Control condition

5. Data

5.1 Participants

In order for the BTS to provide valid and reliable results a large sample size is required (Dimitrov, 2017). This research gathered a total of 400 respondents, 197 respondents in the control condition and 203 respondents in the treatment condition. Across both treatments there were 222 (55.5%) male respondents and 178 (44.5%) female respondents (See Appendix C.1.1). The age of the sample ranged from 18 year until 82 years. The mean age was 25.9 years with a standard deviation of 11.8. Furthermore, the nationality composition consisted of 304 Europeans, 49 Asians, 23 North Americans, 15 South Americans, 6 Africans and 3 Oceanians (See Appendix C.1.1). The majority of the population was single or in a relationship, and very few were married or divorced (See Appendix C.1.1). This relationship distribution can be explained by a fairly young average age. To compare the distribution across conditions a two-sample t-test was preformed for Age (M= 25.91, SD= 11.6, df= 398, t=0.39, p= .700) and Pearson Chi-Squared tests were implemented for Gender (X^2 = 0.29, p= .590), Nationality (X^2 = 3.52, p= .620) and Relationship Status (X^2 = 4.19, p= .520). All of these tests are rejected at the 5%

significance level, displaying the distribution of the two-sample population's demographics to be fairly equal. This removes any bias or effects that may have surfaced from the differences found amongst the treatment and control condition. Consequently, due to this equivalence in the two sample populations, if any difference in the responses occurs it is more likely to have been caused by the different administered surveys rather than external bias.

5.2 Descriptive Statistics

The responses to Q1 (i.e., *"What was the main reason why you lied?"*) and Q2 (i.e., *"Have you lied more than once in the past 24 hours?*") across both conditions were agglomerated and descriptive statistics for both questions were gathered (Table 9-10). Of the total sample population, when answering Q1, 44.25 % reported to have lied for selfless (i.e., selected *"Spare their feelings"*) reasons and 55.75% reported to have lied for selfish reasons (i.e., selected "*Spare yourself from being in an awkward position, conflict or the eventual loss of the person").* Concerning Q2, 71.25% claimed that they did not lie more than once in the past 24 hours and 28.75% claimed the opposite.

Table 9: Q1 (*"What was the main reason why you lied?"*) Distribution

Q1	**Frequency**	**Percent**
"Spare their feelings"	177	44.25
"Spare yourself…"	223	55.75
Total	400	100.00

Table 10: Q2 *("Have you lied more than once in the past 24 hours?)* Distribution

Q2	**Frequency**	**Percent**
"Yes"	115	28.75
"No"	285	71.25
Total	400	100.00

6. Results

6.1 Bayesian Truth Serum validation

In order to confirm that the common prior assumption of the BTS was met, a two-sample t-test for the means obtained from the two percentage estimation questions (i.e., "*"What percentage of them do you feel will provide the same answer as you in the previous question?"*), which followed Q1 (i.e., "*What was the main reason why you lied?")* and Q2 (i.e., *"Have you lied more than once in the past 24 hours?"*). The common prior assumption stipulates that a person, who truthfully selects option 1, for example, will have a tendency to estimate a higher percentage of the population choosing option 1, than a person who chose option 2.

For both percentage estimation questions, the results of the two-sample t-test suggests that people admitting to a certain type of behaviour have a tendency of reporting a significantly higher percentage estimate, than the people who do not admit to the behaviour. The difference between the percentage means for the percentage estimation question, following Q1 is significant at the 5% level (M= 24.99, SD= 21.25, df= 201, t=10.22, p <.01). In other words, a respondent who truthfully reported to have lied for selfish reasons will estimate a higher percentage of the population, which shares the same selfish behaviour, than a person who admitted to have lied for selfless reasons. Similarly, the difference between the

percentage means for the percentage estimation question following Q2 is also significant at the 5% level (M=19.16, SD= 24.01, df= 201, t= 5.62, p <.01). This validates the efficiency of the BTS in this context, as it implies that the truth telling-incentive mechanism preformed well. Consequently, it certifies that the answers obtained from the BTS condition can serve as an appropriate truth base to identify what types of lying behaviours people display. In addition, the answers obtained in the BTS condition will also allow for a reliable comparison with the answers obtained in the control group. Thereof, any significant difference in the responses obtained between the two conditions would suggest that the people in the control group were not being truthful.

6.2 Hypothesis 1 & 2

The first and second hypotheses explore what people's actual lying behaviours are. In order to guarantee that these hypotheses are rejected or accepted correctly, they need to rely on truthful responses provided by participants. Since the BTS functioning was validated in Section 6.1, it implied that the people behaved according to the Bayesian Nash Equilibrium and therefore would indicate that the majority of the answers obtained in the BTS are reliable. Consequently, these two hypotheses will focus solely on the results for Q1 (i.e., *"What was the main reason why you lied?"*) and Q2 (i.e., *"Have you lied more than once in the past 24 hours?"*) obtained in the BTS condition. Both Q1 and Q2 offer two-answer options, of which the respondents can only select one. To see which answer for Q1 and Q2 was most frequently selected a proportion test of the answer options of Q1 and Q2 was implemented.

The first hypothesis was centered on whether people were more selfish or selfless when telling altruistic lies. In other words, was option 1 ("Spare their feelings") or option 2 ("*Spare yourself from being in an awkward position, conflict or the eventual loss of the person"*) more common? It was gathered that 89 (43.84%) respondents chose option 1 and 114 (56.16%) respondents chose option 2. The test proportion shows that the amount of people admitting to lie for selfish reasons is significantly

greater than the amount of people declaring to have lied solely for the benefit of the other person, at the 5% level (M= 0.12, SE= 0.05, z=2.48, p= .01). This confirms H1, that on average people have more selfish motives when telling altruistic lies.

H1: *On average people's real motives behind altruistic lies are for more selfish than selfless*

The second hypothesis (H2) predicted that people would report having lied more than once in the past 24 hours. Hence, it H4 was based on Q2. In the BTS condition, 62 (30.54%) respondents chose option 1 (i.e., "Yes") and 141 (69.46%) respondents chose option 2 (i.e., No). The output of the proportion test demonstrates that the frequency of option 1 being selected was significantly smaller than the frequency of option 2, being chosen, at the 5% level (M=0.39, SE= 0.05, z= 7.84, p < .01). Therefore, from this result it would seem that the majority of people reported, to have not lied more than once in a time frame of 24 hours. This rejects H4.

H2: *On average people lie more than once in 24 hours*

6.3 Hypothesis 3 & 4

In order to test the third and fourth hypotheses and observe whether people in the control condition provided more socially desirable responses than people in the BTS condition for Q1 and Q2, Pearson Chi-Squared Tests were preformed for each question.

The third hypothesis (H3) examines whether people lied about the true motives behind their altruistic lies. This links the first hypothesis directly to Q1 (i.e., "*What was the main reason why you lied?"*). When answering Q1 respondents had the choice between two options: *"Spare their feelings"* or "*Spare yourself from being in an awkward position, conflict or the eventual loss of the person"*. In the control

condition, 88 respondents chose *"Spare their feelings"* and 109 chose "*Spare yourself from being in an awkward position, conflict or the eventual loss of the person"*. Similarly, in the BTS condition, 89 people chose *"Spare their feelings"* and 114 chose "*Spare yourself from being in an awkward position, conflict or the eventual loss of the person"*. See Figure 2 below. The result of the Chi-Squared test demonstrate that the difference in the responses for Q1 between the two conditions is insignificant at the 5% level, X^2 = 0.03, p= 0.87, leading to a rejection H3.

***H3**: On average, participants in the BTS condition will report to have lied for selfish reasons, more times than participants in the Control condition.*

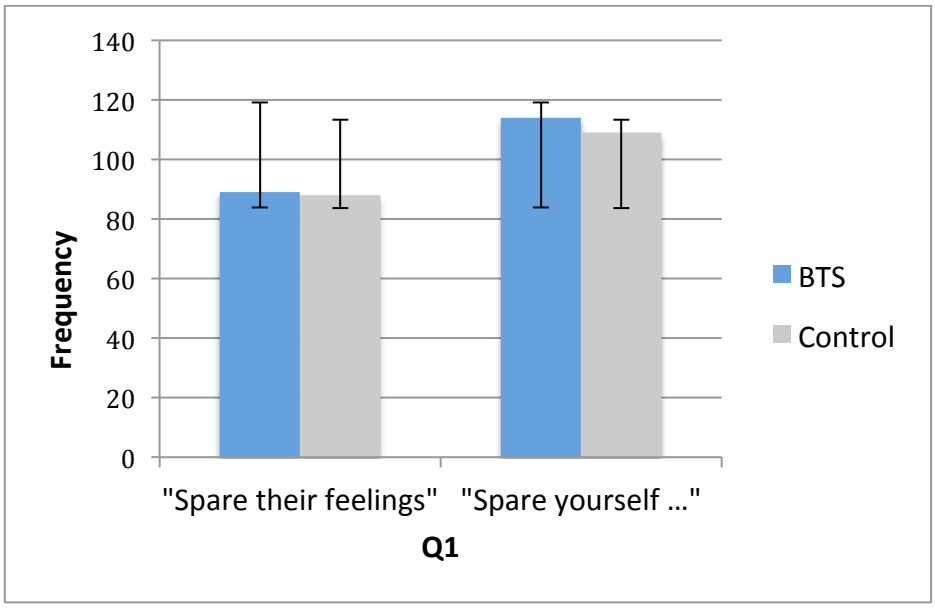


Figure 2: Q1 *("What was the main reason why you lied?"*) answer option distribution across conditions standard deviation error-bars

The fourth hypothesis (H4) focuses on whether people lie about their lie frequency in the past 24 hours. Consequently, Q2 (*"Have you lied more than once in the past 24 hours?"*) served to confirm or reject H4. In the control condition, 53 respondents answered *'Yes'* (i.e., to have lied more than once in the past 24 hours) and 144 answered *'No'* (i.e., to have not lied more than once in the past 24 hours). Likewise, in the BTS condition 62 participants responded *'Yes'* and 141 participants selected *'No'*. See Figure 3 below. The reported Chi-Squared test result, proves that no

significant difference, in the responses reported for Q2, was found between both conditions, at the 5% significance level, X^2 = 0.65, p= 0.42. This rejects H4.

H4: *On average, participants in the BTS condition will report to have lied more than once in the past 24 hours, more times than participants in the Control condition*

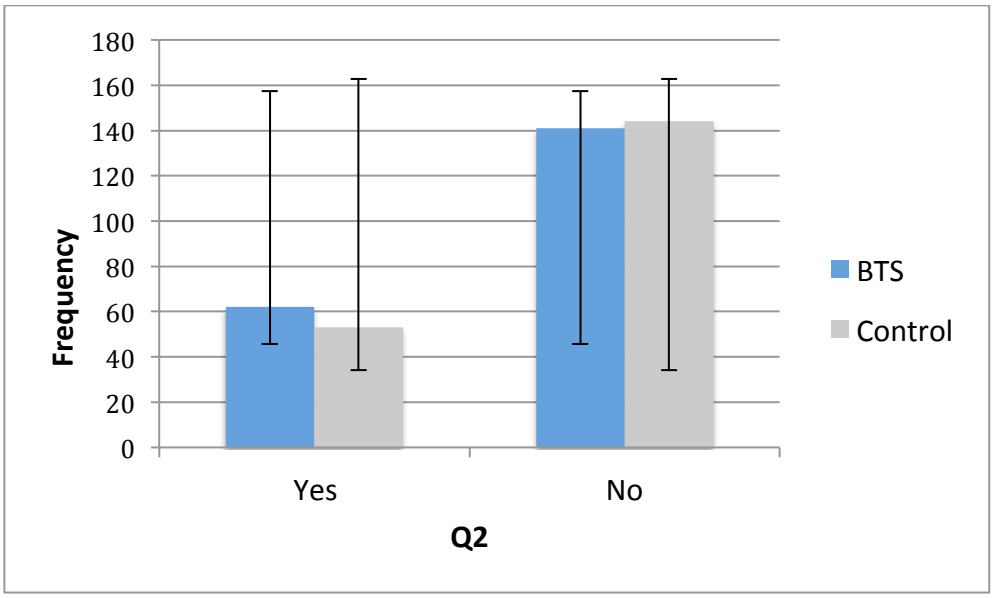


Figure 3: Q2 *("Have you lied more than once in the past 24 hours?")* answer option distribution across conditions and standard deviation error-bars

7. Discussion & Limitations

7.1 General Discussion

The predominant purpose of this study was to use the BTS in order to test whether people from the control condition lied more often, when answering their survey, in comparison to the people in the BTS condition. Two different aspects of lies were investigated: the motivations behind altruistic lies and people's lie frequency. The first main question (i.e., Q1: *"What was the main reason why you lied"*) focused on altruistic lies and more precisely tried to understand what the main motivation people had in telling such lies. Moreover, the second main question (i.e., Q2: *"Have*

you lie more than once in the past 24 hours?") asked if the respondent had lied more than once in the past 24 hours in order to analyze respondents' average lie frequency. The answers to Q1 and Q2, collected from the control condition were not significantly different than the answers obtained from the BTS condition. These results suggest that either people in the control condition were as honest as people in the BTS condition or that the truth telling incentive in the BTS condition did not work, leading people in the BTS to report dishonestly. However, the incident that people in the BTS were in fact being dishonest is extremely unlikely, since if this were the case it would be expected that people would provide more socially desirable responses. For instance, for the first question that asked about the main motive behind people's altruistic lies, if respondents were being dishonest it would be more plausible that they would report to have lied for purely altruistic purposes, than to report to have lied for selfish purposes. Through the two-sample t-test, it was shown that the majority of people in the BTS condition reported to have lied for more selfish purposes than altruistic purposes. However, it is true that for the second question, which asked if people lied more than once in the past 24 hours, the answer that was most common (i.e., "No"), was also the most socially desirable answer. Consequently, from this perspective in could be argued that people in the BTS condition may have been as dishonest as people in the control condition. Nevertheless, the results obtained for the t-test that confirmed the common prior assumption of the BTS was met, signal that people behaved according the Bayesian Nash Equilibrium, which is to tell the truth. This validates the performance of the BTS to the types of questions applied in this investigation. The findings obtained for the common prior assumption are in line with other papers that have also validated the functioning of the BTS (Prelec, 2004; Weaver et al., 2013; Abolmagd, 2017; Baillon, 2017; Cvitanić et al., n.d.). All of these arguments together support the explanation that that the respondents in control condition were being as honest as those in the BTS condition and not the contrary claiming that the respondents from the BTS condition were being as dishonest as those in the control condition. Therefore, this justifies the use of the BTS to get an understanding of people's self-reported lying behaviours by observing the answers obtained for Q1 and Q2, in the BTS condition.

Fisher (1993) noted that people are often discouraged to reveal the truth about their questionable behaviours. Since, this study was based on questionable behaviour (i.e., lying), there was a chance for the participants to report dishonest answers, and thus cause the answers of the respondents to suffer from social desirability bias (Fisher & Katz, 2000). Other papers, investigating into people's lying behaviours and, similarly to this experiment, being based on self-reports, are believed to have suffered from under-reporting (DePaulo, 1996; Serota et al., 2010; Debey et al., 2010). However, unlike these studies, the current experiment made use of a truth serum (i.e., the BTS), which demonstrated to function and preform correctly. As the results of the experiment show that the answers between control condition and the BTS condition were not significantly different, it suggests that people in the control condition did not under-report compared to the people who were faced with the truth serum (i.e., BTS sample). In other words, it seems that people in the control condition were not inclined to lie because of any potential social unease they may have felt, when having to report their lying behaviours. This is also confirmed because for the first question (i.e., *"What was the **main** reason why you lied?"*), the majority of people in the control condition did not choose the socially desirable answer. That is, they admitted to have lied for more selfish reasons, which would normally constitute a questionable behaviour.

However, it cannot be generalized that since the control condition in this experiment did not suffer from under-reporting, that this holds for all self-report surveys. Perhaps, participants in the control condition may have been compelled to tell the truth by the way in which the instructions were laid out as well as how the questions were formulated, in the experiment. That is, the questions reinforced the aspect of honesty, which could have also served as a 'truth serum', maneuvering people towards answering truthfully. For instance, the instruction *"please answer **truthfully**"* by highlighting and putting an accent on the word 'truthfully' could push respondents in being more conscious about their answers and hence take more time to reflect on them. Although, DePaulo (1996) emphasized the aspect of "accuracy and conscientiousness" to the respondents in her experiment, nevertheless she did not explicitly highlight the aspect of honesty. Similarly, Serota et al. (2010) and

Debey et al. (2010) were careful to provide a morally neutral explanation to lying, but just like DePaulo (1996) focused more on the lying than the honesty feature. That is in the explanations and introduction phase provided in their study they did not put an emphasis on how important it was for the respondents to report honestly. Contrary to the current study, which went as far as stressing the word ***'truthfully'***. Future research could be done to test whether asking respondents to answer honestly would provide different answers, than if it were not included. To test this, three conditions would be needed: one control condition not including this honesty aspect, one condition including the honesty aspect and another condition including the BTS and the honesty aspect.

Another possible explanation, as to why reinforcing the honesty feature could have served as a truth serum is a more psychological occurence. Social psychology has shown that when people are asked to take an oath, the chances of being dishonest is significantly diminished (Kiesler, 1971). This is caused because people view it as making a commitment and are reminded about their moral stadards. Jaquemet, Joule, Luchini, and Shogre (2013) investigated the effect of making respondents sign a solemn oath agreement prior to responding to a questionnaire. Likewise, Mazar et al. (2008) implemented the same concept of reminding people about their moral code, by asking people to write down the Ten Commandments prior to undertaking a task where dishonesty could increase the payoffs of the respondents. Both studies discovered that making people reminding people of their moral standards, creates a psychological commitment in the mind of the respondent, which serves as effective truth serum. Confronting respondents to instructions and questions clearly reinforcing the aspect of honesty, could have served as an unconscious commitment device and reminder of their morality, where respondents felt compelled to answer truthfully.

In general, a possible cause for unreliable questionnaire results may be a lack of attention, especially when answering long and complex questionnaires. Some lies may not be cognitively registered by people, and in combination with a lack of focus when answering a questionnaire, can cause participants to underreport. The BTS'

main purpose is to avoid the latter phenomenon, by creating an incentive for respondents to invest themselves more into surveys as well as asking respondents to reflect on the behaviour other respondents may have, to enhance the quality of the data collected (Prelec, 2004). Survey length is an essential factor affecting the quality of the data collected. The longer the survey and the more the questions become complex, the more the quality obtained from the survey will decrease. Since the survey conducted for this research consisted of only two short and simple questions, the respondents were most likely encouraged to provide mindful and truthful responses (Galesic & Bosnjak, 2009). Therefore, this could be another reason why the results obtained from the control condition were not significantly different from the ones obtained in the BTS condition. In addition, respondents seemed to be cognitively invested when asking people to complete the survey in real life. The majority of the subjects, once confronted with the main questions, would take more time in answering them. Perhaps, if the experiment contained more complex questions and its duration was increased, then a divergence in the responses from the control condition, compared to the responses from the BTS condition, may have been more apparent. That is, people in the control condition would have been less invested in answering a complex questionnaire compared to the participants from the BTS condition, since in the latter the incentive was tied to the type of responses offered. Future research could potentially be conducted on whether the manner in which questions are formulated and presented may influence the responses of people and conceivably serve as truth serums. If these factors concerning questionnaire length, presentation, and question formulation would influence people in providing truthful responses reducing then the use of the BTS in this experiment would not be justified, as the control condition provided the same results on average. Thus, the results of both conditions could have been agglomerated into one, in order to get a better understanding of people's lying behaviours, instead of only observing the results obtained from the BTS condition. Thereof, a test of proportion of the frequency of each answer option for Q1 and Q2 being reported, would be implemented.

The evidence, acquired from this experiment, that people had more selfish motivations when telling altruistic lies, are aligned with Hobbe's (1651) statement: *"no man giveth but with intention of good to himself"*. This suggests that people have a tendency to justify that they are lying for an altruistic purpose when in fact they are lying primarily out of convenience for themselves. These findings would support Shalve et al.'s (2015) pre- and post-violation justification schematic (Appendix A.2) that demonstrates how a person can internally justify wrongdoings and still feel moral. In their paper they highlighted how "lies causing no harm to a concrete victim but benefiting concrete others also serve as pre-violation justifications" (Shalvi et al., 2015). This human need to feel moral can be explained by Dufwenberg and Gneezy's (2000) idea of how human beings experience a disutility when acting immoral. This phenomenon seems to be a plausible explanation to the results obtained in this study. Indeed, people may be pushed to justify their lies as being purely altruistic, when in fact the main motivation is to serve the self, just to feel good about themselves. In general, people have the need to feel good about their persona, leading them to even lie to themselves (Zerbe & Paulhus, 1987).

Moreover, Erat and Gneezy (2012) labeled lies that both help others and the liar as *Pareto white lies* and defined altruistic white lies as "a lie that can harm the liar but help the other person". Therefore, according to Erat and Gneezy's (2012) definition, the results obtained in this study would indicate that the majority of people in the BTS condition, report to tell more *Pareto white lies*. Another interesting point was how easy it was to have people admit to their questionable behaviour. Another possible explanation may be that in general, without much consideration a person may label his/her lie as an altruistic lie, when one of the reasons appears to benefit the other person amongst many others, as this would avoid any internal immoral turmoil or cognitive dissonance that the person would have to deal with, if they would face the truth (Festinger, 1957). It is only when the person is asked to think hard what the ***main*** reason is for telling such a lie, that they recognize the ulterior selfish motive.

The findings concerning the lie frequency suggest that people do not lie more than once in 24 hours. More precisely, 60.46% in the BTS condition neglected to have lied more than once in the past 24 hours. It would seem that people, in a day, lie at most once, if not less. DePaulo et al. (1996) obtained that one average people tell one to two lies in a day. More specifically, they obtained that college students, who were on average 18.69 years old, reported to lie two times in a day and that the community, who were on average 34.19 years old, reported to lie once in a day. Focusing on the findings obtained from college students, who reported an average of 1.96 lie in a day, would seem to oppose what was obtained in this experiment. However, concerning the community, who's mean number of lies was 0. 97, it would seem to be more in line with the results of the current investigation. Unfortunately, due to the way in which the question (i.e., *"Have you lied more than once in the past 24 hours?"*) was asked in the current survey, we can not know the exact number of lies told by the participants reporting have not lied more than once in the past 24 hours.

Serota et al. (2010) found out of all the participants they gathered, 60% told no lies at all and almost half of the lies reported in the study, were told by only 5% of all the participants. Once again, due to the way in which the question phrased, it cannot be said that the results obtained oppose those of Serota et al. (2010). This is because, we do not know if the respondents neglecting to have lied more than once in the past 24 hours, lied at least once or did not lie at all. Nevertheless, Serota et al.'s (2010) study does agree with the statement that the majority of people do not seem to lie more than once in 24 hours. Likewise, in the sample population of this investigation we notice a minority (28.75%) who report to lie more than once in 24 hours which is in line with the non-normal distribution of the lies reported by Serota et al's (2010) and Debey et al. (2010). Specifically, they found a minority of prolific liars told the majority of the lies. Hence, lying to this minority seems to be more common compared to the rest of the sample. Anecdotally, when gathering data in person, contrasting behaviours amongst participants were obtained. On the one hand, some participants seemed to really think hard whether they had lied more than once in the past 24 hours and would take a few moments to reflect. Whilst on

the other hand, some respondents laughed at the question and without hesitation knew they had lied more than once. Therefore, just like Serota et al. (2010) reported, it is possible that DePaulo and al.'s (1996) results for the college students which acquired an average of 1.96 lies in day, was driven by a minority of prolific liars, causing the average frequency of lies to increase. From the findings of this experiment, it is suggested that on average the majority of people report to lie at most once in a time lapse of 24 hours.

7.1 Limitations

One of the main limitations of this study concerned the monetary incentive provided for the BTS condition. Only one participant could win the prize (i.e., €20) if they obtained the highest BTS score (i.e., there answers were the most truthful). Therefore, the truth incentive may not have been as effective and efficient, than if each person would obtain a prize based on the score they obtained (Dimitrov, 2017).

Moreover, the way in which the questions were formulated may have directed people's answers. Q1 (i.e., *"What was the main reason why you lied?*) offered two possible answer options: (1) *Spare their feelings* and (2) *Spare yourself from being in an awkward position, conflict or the eventual loss of the person*. Since the instructions preceding Q1 offered a detailed explanation of how an altruistic lie can have more selfish reasons, it may have framed people into choosing option (2). More precisely, by putting an emphasis on how an altruistic lie can have selfish undertones, may have neutralized the act as common. Therefore, as a result people may have felt less shame in admitting it, since from the instructions provided it would appear that telling an 'altruistic' lie for selfish motives is a prevailing occurrence. Similarly, seeing as telling an 'altruistic' lie for selfish motives is common may have reduced the ethical dissonance people may have felt. This is in line with Barkan, Ayal, Gino, and Ariely's (2012) findings that showed how people reduce ethical dissonance by judging others more harshly or at least validating their immoral behaviour through the immoral acts of other people. Perhaps, if less explanation was provided concerning altruistic lies, then people in the control group

may have been more compelled to lie, and thus a significant difference between the answers in the two conditions would have resulted. However, this may have put into jeopardy the understanding of what constitutes an altruistic lie.

Finally, another limitation that may have contributed to the lack in difference between the two conditions was the insistence on the honesty aspect throughout the survey. Similarly, to the examples of how an altruistic lie can hide more selfish motives, this honesty emphasis may have influenced people into providing the truth and therefore served as a truth serum in itself.

8. Conclusion

In conclusion, the main finding gathered on peoples' altruistic lies was that people report to lying for more selfish purposes than purely lying for the benefit of the other person. In addition, the majority of the respondents in the BTS condition reported to not lie more than once in the past 24 hours. The objective of this study was to compare the responses obtained to lying behaviour questions from two different conditions: a control condition and BTS condition. Although no significant difference was found between the people undergoing the control condition and the ones undergoing the BTS (i.e., lying was not detected amongst the people in the control condition), nevertheless this does not undermine the performance of the BTS. Throughout this experiments as well as in the literature (Abolmagd, 2017; Dimitrov, 2017; Schneider, 2017), the results obtained showed that the BTS functions well when the truth is unverifiable. This supports the validity of the BTS and its usage in a setting in which lying behaviour is the topic of interest.

9. References

Abolmagd, M.I.M.T. (2017). *Tell me the Unverifiable Truth - Bayesian Truth Serum and Social Desirability Bias*. *Economics*. Retrieved from: http://hdl.handle.net/2105/37833

Ariely, D., & Melamede Y. (2015). *Dis(honesty)* [Video file]. Retrieved from: https://www.youtube.com/watch?v=RVix6vognrY

Arnold, H. J., & Feldman, D. C. (1981). Social desirability response bias in self-report choice situations. *Academy of Management Journal*, *24*(2), 377-385.

Augustine. *Treaties on various issues*. Catholic University of America Press; Washington, DC: 1952.

Backbier, E., Hoogstraten, J., & Terwogt-Kouwenhoven, K. M. (1997). Situational Determinants of the Acceptability of Telling Lies 1. *Journal of Applied Social Psychology*, *27*(12), 1048-1062.

Baillon, A. (2017). Bayesian markets to elicit private information. *Proceedings of the National Academy of Sciences*, *114*(30), 7958-7962.

Barasch, A., Levine, E. E., Berman, J. Z., & Small, D. A. (2014). Selfish or selfless? On the signal value of emotion in altruistic behavior. *Journal of personality and social psychology*, *107*(3), 393.

Batson, C. D. (1991). *The Altruism Question: Toward a Social-Psychological Answer*. Hillsdale, NJ: Erlbaum.

Bergstrom, C. T. (2009). Dealing with deception in biology. In: B. Harrington (ed.), *Deception: From ancient empires to internet dating*, 19-37. Stanford, CA: Stanford University Press.

Bok, S. (1999). *Lying: Moral choice in public and private life*. Vintage.

Bond Jr, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and social psychology Review*, *10*(3), 214-234.

Bryant, E. M. (2008). Real lies, white lies and gray lies: Towards a typology of deception. *Kaleidoscope: A Graduate Journal of Qualitative Communication Research*, *7*, 23.

Burkey, L. M. (1965). The case against the polygraph. *American Bar Association*

Journal, 855-857.

Cvitanić, J., Prelec, D., Radas, S., & Šikić, H. (n.d) Bayesian Truth Serum and Information Theory: Game of Duels. Retrieved from https://web.math.pmf.unizg.hr/zzvs/userfiles/downloads/truth_serum.pdf

Debey, E., De Schryver, M., Logan, G. D., Suchotzki, K., & Verschuere, B. (2015). From junior to senior Pinocchio: A cross-sectional lifespan investigation of deception. *Acta psychologica*, *160*, 58-68.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of personality and social psychology*, *70*(5), 979.

Dufwenberg, M., & Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and economic Behavior*, *30*(2), 163-182.

Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, *58*(4), 723-733.

G.B. Dimitrov. (2017, August 29). *Bayesian Truth Serum Fused Conjoint*. *Economics*. Retrieved from http://hdl.handle.net/2105/39589

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

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

Fisher, R. J., & Katz, J. E. (2000). Social-desirability bias and the validity of self-reported values. *Psychology & Marketing, 17*(2), 105-120.

Ford, E. B. (2006). Lie detection: Historical, neuropsychiatric and legal dimensions. *International Journal of Law and Psychiatry*, *29*(3), 159-177.

Fu, G., Xu, F., Cameron, C. A., Heyman, G., & Lee, K. (2007). Cross-cultural differences in children's choices, categorizations, and evaluations of truths and lies. *Developmental Psychology*, *43*(2), 278.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public opinion quarterly*, *73*(2), 349-360.

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

Hubbard, R., & Little, E. L. (1988). Promised contributions to charity and mail survey responses: Replication with extension. *Public Opinion Quarterly*, *52*(2), 223-230.

International Assoc of Chiefs of Police. (1975). PSYCHOLOGICAL STRESS EVALUATOR. *PWC BULLETIN*, *74*(12), 6.

Jaquemet, N., Joule, R. V., Luchini, S., & Shogren, J. F. (2013). Preference elicitation under oath. *Journal of Environmental Economics and Management*, *65*(1), 110-132.

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

Kant, I. On a supposed right to lie fromaltruistic motives. In: Beck, LW., editor. Critical of practical reason and other writings. University of Chicago Press; Chicago: 1949. p. 346-350.

Kiesler, C. A. (1971). *The psychology of commitment: Experiments linking behavior to belief*. Academic Press.

Kukla-Gryz, A., Tyrowicz, J., Krawczyk, M., & Siwiński, K. (2015). We all do it, but are we willing to admit? Incentivizing digital pirates' confessions. *Applied Economics Letters*, *22*(3), 184-188.

Kullback, S. (1987). Letter to the editor: The Kullback-Leibler distance.

Larson, J. A. (1921). Modification of the Marston deception test. *J. Am. Inst. Crim. L. & Criminology*, *12*, 390.

Levine, T. R., Ali, M. V., Dean, M., Abdulla, R. A., & Garcia-Ruano, K. (2016). Toward a pan-cultural typology of deception motives. *Journal of Intercultural communIcatIon research*, *45*(1), 1-12.

Loughran, T. A., Paternoster, R., & Thomas, K. J. (2014). Incentivizing responses to self-report questions in perceptual deterrence studies: An investigation of the validity of deterrence theory using Bayesian truth serum. *Journal of Quantitative Criminology*, *30*(4), 677-707.

Marston, W. M. (1915). William Moulton Marston. Retrieved from: http://www.popflock.com/learn?s=William_Moulton_Marston

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.

Miller, S. R., Bailey, B. P., & Kirlik, A. (2014). Exploring the utility of Bayesian truth serum for assessing design knowledge. *Human–Computer Interaction*, *29*(5-6), 487-515.

Morris, R. (2017, May 18). Why We Lie: The Science Behind Our Deceptive Ways. Retrieved March 16, 2018, from https://www.nationalgeographic.com/magazine/2017/06/lying-hoax-false-fibs-science/

Nyberg, D. (1993). *The varnished truth: Truth telling and deceiving in ordinary life*. University of Chicago Press.

Palmiotto, M. J. (1983). An historical review of lie-detection methods used in detecting criminal acts. *Canadian Police College Journal*, *7*(3), 206-216.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of experimental social psychology*, *13*(3), 279-301.

Schneider, L.S. (2017, July 20). *A Non-Monetary Truth-Telling Incentive*. *Economics*. Retrieved from http://hdl.handle.net/2105/39426

Serota, K. B. (2015). *Kim Serota: Why people lie* [Video file]. Retrieved from: https://www.ted.com/tedx/events/14788

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America: Three studies of self-reported lies. *Human Communication Research*, *36*(1), 2-25.

Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science*, *24*(2), 125-130.

Tenbrunsel, A. E., & Messick, D. M. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social justice research*, *17*(2), 223-236.

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

Vrij, A. (2000). Detecting Lies and Deceit: The Psychology of Lying and Implications for Professional Practice.(Wiley Series on the Psychology of Crime, Policing and Law).

Vrij, A. (2007). Deception: A social lubricant and a selfish act. *Social communication*, 309-342.

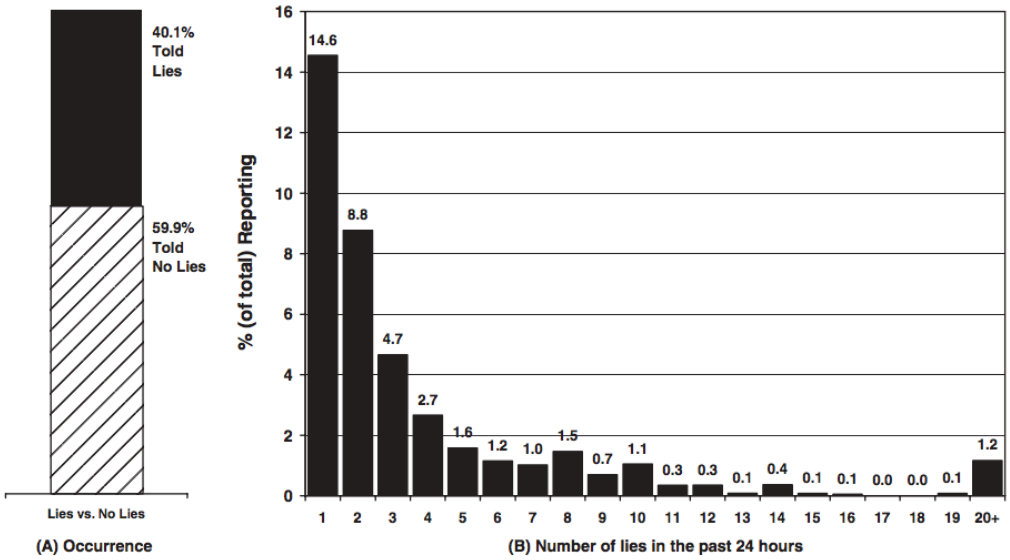
Warriner, K., Goyder, J., Gjertsen, H., Hohner, P., & McSpurren, K. (1996). Charities, No; Lotteries, No; Cash, Yes: Main Effects and Interactions in a Canadian Incentives Experiment. *The Public Opinion Quarterly, 60*(4), 542-562. Retrieved from http://www.jstor.org.eur.idm.oclc.org/stable/2749634

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

Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Review*, *12*(2), 250-264.

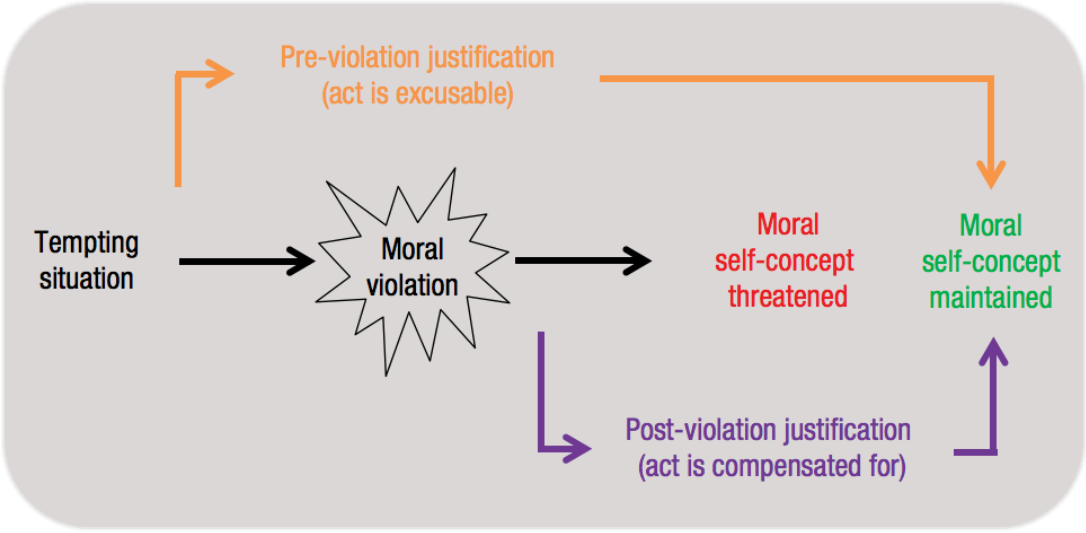
10. Appendix

A.1. Lie Frequency Distribution



Source: Serota et al. (2010)

A.2. Pre- and Post-Violation Justification Schematic



Source: Shalvi et al. (2015)

B.1. Experiment

B.1.1. Introduction

Hi, my name is Amielle and I am currently investigating into people's lying behaviours. For instance, the main reason why people lie. This survey will not take you more than **5 minutes** to answer and in return you may win **€ 20**.

Your data will be stored only for the purpose of this research. If you wish to get the chance of winning the 20 euros you will need to provide your email address in order to contact you in the future. Once the prize has been delivered all the email addresses will be deleted.

However, if you wish to remain **completely** anonymous this is also possible but in this case you will not be able to win the 20 euro prize.
*(If you have any questions/complaints please email: **amielle.guilloux@gmail.com**)*

By clicking **next** you consent that you are 18 years of age or older and that your data will be collected, stored and used for this research.

B.1.2. Demographic questions

How old are you?

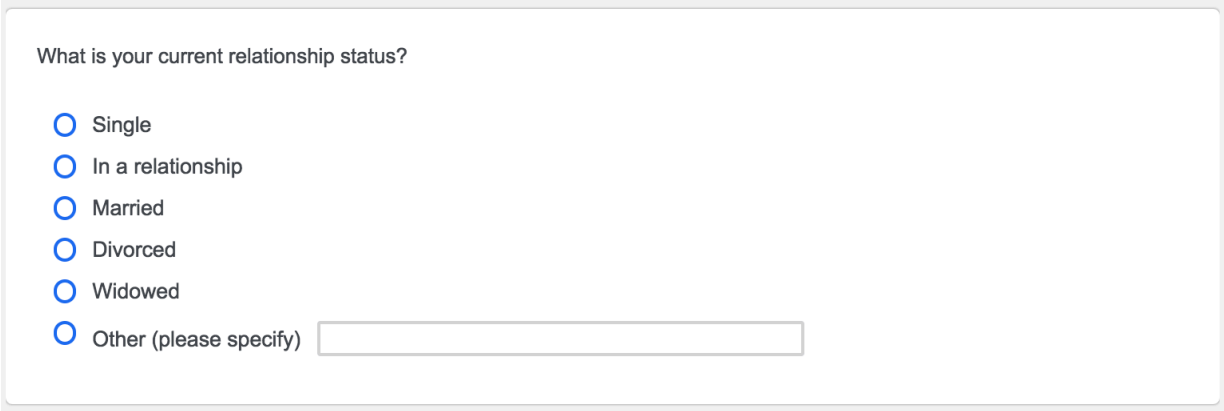
What is your gender?

- ○ Male
- ○ Female

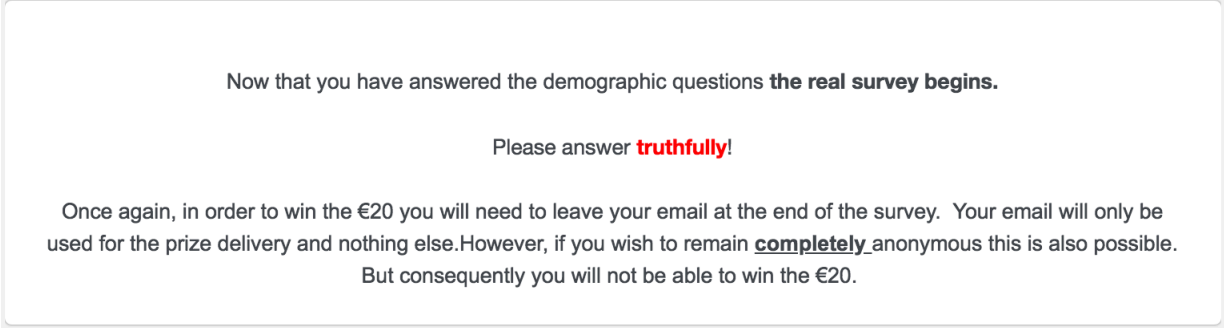
What is your nationality?

*(Note: If you have more than one then what is the nationality you identify **most** to?)*

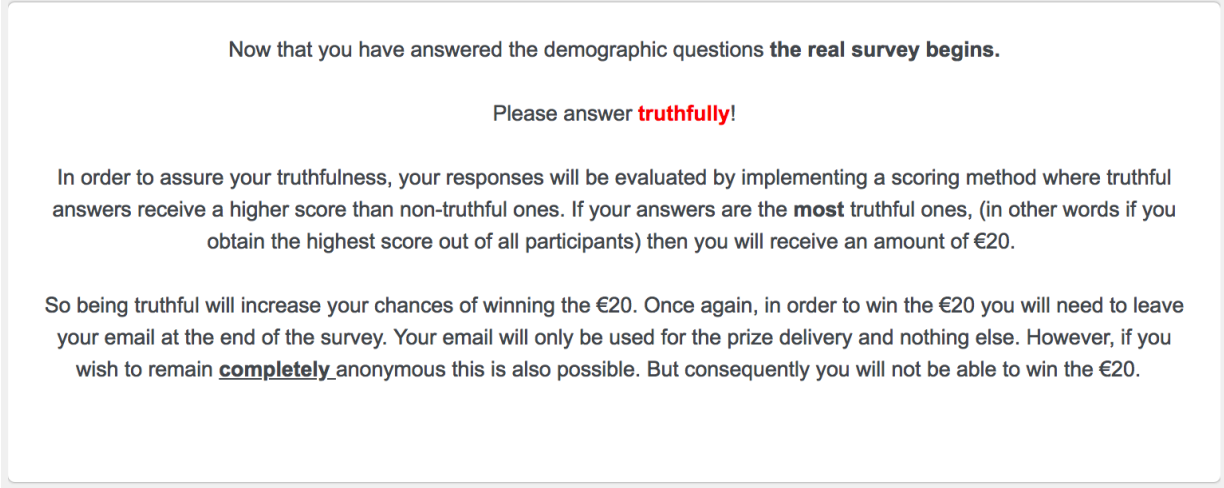
- ○ European
- ○ Asian
- ○ African
- ○ North American
- ○ South American
- ○ Oceanian



B.1.3. Control group instructions



B.1.4. BTS instructions



B.1.5. Control group survey

Please **read** the instructions before answering the questions.

What is a lie?
A lie is any form of **false-hood** told by a person. It can be an exaggeration just like it can also be an omission of the truth or simply a false statement.

The reasons why we lie depend on variety of things. For instance, we can lie to embellish our image, to protect ourselves or even for the benefit of others. This survey will focus on the latter i.e. the type of lies we tell for the benefit of others.

Examples of lies to protect another person:

- Telling a person "those pants look great" when they don't in order to spare their feelings
- Not telling your partner you are cheating in order to avoid them getting hurt
- Exaggerating on how good a person looks to boost their confidence

Have you ever lied to spare someone's feelings?

- ○ Yes
- ○ No

Is there an other reason why we tell social lies? Is the **main** reason why we tell such lies for the **benefit** of the other person or is there a **more egoistical** reason?

For example:

- Telling a person "those pants look great" when they don't in order to spare their feelings **or to avoid an awkward situation**
- Not telling your partner you are cheating in order to avoid them getting hurt **or to avoid losing your partner**
- Exaggerating on how good a person looks to boost their confidence **or in order to be liked by the person**

Keep the example that came in mind when answering the previous question.

If you are completely **honest**, what was the **main** reason why you lied ?

To:

- ○ Spare their feelings.
- ○ Spare yourself from being in an awkward position, conflict or the eventual loss of this person.

Did you lie more than once in the last 24 hours?

- ○ Yes
- ○ No

B.1.6. BTS survey

Please **read** the instructions before answering the questions.

What is a lie?
A lie is any form of **false-hood** told by a person. It can be an exaggeration just like it can also be an omission of the truth or simply a false statement.

The reasons why we lie depend on variety of things. For instance, we can lie to embellish our image, to protect ourselves or even for the benefit of others. This survey will focus on the latter i.e. the type of lies we tell for the benefit of others.

Examples of lies to protect another person:

- Telling a person "those pants look great" when they don't in order to spare their feelings
- Not telling your partner you are cheating in order to avoid them getting hurt
- Exaggerating on how good a person looks to boost their confidence

Have you ever lied to spare someone's feelings?

- ○ Yes
- ○ No

Is there an other reason why we tell social lies? Is the **main** reason why we tell such lies for the **benefit** of the other person or is there a **more egoistical** reason?

For example:

- Telling a person "those pants look great" when they don't in order to spare their feelings **or to avoid an awkward situation**
- Not telling your partner you are cheating in order to avoid them getting hurt **or to avoid losing your partner**
- Exaggerating on how good a person looks to boost their confidence **or in order to be liked by the person**

Keep the example that came in mind when answering the previous question.

If you are completely **honest**, what was the **main** reason why you lied ?

To:

- ○ Spare their feelings.
- ○ Spare yourself from being in an awkward position, conflict or the eventual loss of this person.

Think about the other respondents answering this survey.

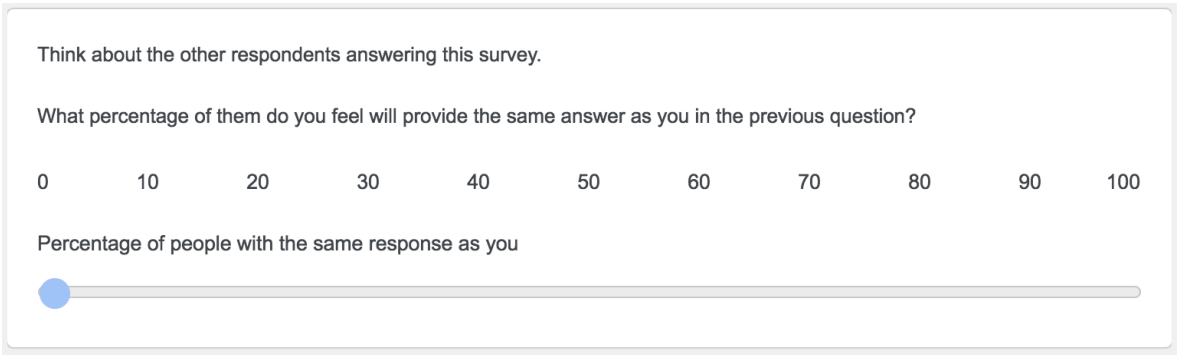
What percentage of them do you feel will provide the same answer as you in the previous question?

0 10 20 30 40 50 60 70 80 90 100

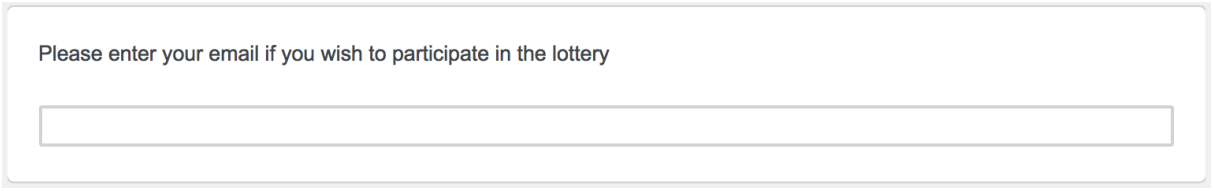
Percentage of people with the same response as you

Did you lie more than once in the last 24 hours?

- ○ Yes
- ○ No



B.1.7. Invitation email



B.1.8. End of survey

C.1: Data

C.1.1. Descriptive statistics

Table 1: Gender distribution

Control

Gender	**0**	**1**	TOTAL
Female	85	98	178
Male	112	110	222
TOTAL	197	203	400

Table 2: Nationality Distribution

Control

Nationality	**0**	**1**	TOTAL
European	151	153	304
Asian	22	27	49
North American	14	9	23
South American	5	10	15
African	3	3	6
Oceanian	2	1	3
TOTAL	197	203	400

Table 3: Relationship Status Distribution

Control

Relationship Status	**0**	**1**	TOTAL
Single	109	104	213
In a Relationship	68	81	149
Married	18	17	35
Divorced	1	0	1
Other	1	1	2
TOTAL	197	203	400

Table 4: Age Distribution

Control

Age	**0**	**1**	**Treatment + Control**
Mean	26.13	25.68	25.9
Standard Dev.	12.18	11.35	11.75

Table 5: Control Distribution

Control	**Frequency**	**Percent**
0	197	49.25
1	203	50.75
Total	400	100.00

Table 6: Gender Distribution in Control and Treatment

Gender	**Frequency**	**Percent**	**Cum.**
Female	178	44.50	44.50
Male	222	55.50	100.00
Total	400	100.00	

Table 7: Relationship Status Distribution in Control and Treatment

Relationship Status	**Frequency**	**Percent**	**Cum.**
Single	213	53.25	53.25
In a Relationship	149	37.25	90.50
Married	35	8.75	99.25
Divorced	1	0.25	99.50
Other	2	0.50	100.00
Total	400	100.00	

Table 8: Nationality Distribution in Control and Treatment

Nationality	**Frequency**	**Percent**	**Cum.**
European	304	76.00	76.00
Asian	49	12.25	88.25
North American	23	5.75	94.00
South American	15	3.75	97.75
African	6	1.50	99.25
Oceanian	3	0.75	100.00
Total	400	100.00	