ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis

Business and Economics – Behavioural Economics

Bayesian Truth Serum and Upward Feedback

Can you elicit truthfulness from your employees' subjective opinions?

Author: Hai Binh Dang

Student ID: 534159hd

Supervisor: Tong Wang

Second assessor: Aurelien Baillon

August 9th, 2021

Abstract

In light of concerns about the quality of employees' upward feedback and potential organisational challenges pertaining to the dishonesty of feedback, this study was conducted to examine the benefits of the Bayesian Truth Serum (BTS). A review of existing literature revealed that the BTS has been proven to work as an effective instrument for improving the likelihood of truth acquisition in some social contexts. Meanwhile, upward feedback remains ineffective in playing its functions. The hindrance derives from cognitive biases in the process of feedback provision, which distort genuine and honest responses from employees. Provided that no study by far has explored the application of the BTS in upward feedback improvement, a confirmatory study was performed to testify the benefits of the BTS-design, especially mitigating biases, through a web-based BTS-integrated survey, as well as a mechanism of reward for truthful answers. As the power test confirmed the validity of the sample size in providing statistically meaningful results, quantitative analytic methods were performed with two objectives: to test whether the BTS design is statistically effective in eliciting honest feedback; and to test whether the intervention of the BTS can mitigate the effect of different biases in providing positive, presumably dishonest, responses in upward feedback. The outputs implied that the BTS design essentially induces honest answers in upward feedback, as the likelihood of positive responses in the treatment group was considerably lower than that of the counterpart. The integration of the BTS in designing employee surveys would work to a certain extent, namely leniency, cynicism and relation bias mitigation, yet it needs more measures to prevent fear for retaliation.

Keywords: Bayesian Truth Serum, upward feedback, truth-telling incentive, behavioural bias

Contents

Introduction 4

Literature Review 6

Bayesian Truth Serum *6*

Upward Feedback *10*

Experimental Design 16

Experimental Strategy *16*

Data Collection *16*

Hypotheses *19*

Ethic Consideration *20*

Data Analysis 20

Data Description *20*

Descriptive Statistics *21*

Correlation Analysis *23*

Results 25

Prevalence Estimates *25*

Impacts of the BTS Application in Moderating Relationship between Cognitive Biases and Truthfulness of Upward Feedback *25*

Discussion 27

Practical Implications *29*

Limitations and Future Study Suggestions *31*

Conclusion 32

References 34

Appendices 38

Introduction

Upward communication, known as transference of information and meaning inference from a sender in a lower organisational hierarchy to a receiver in a higher hierarchy, is considered vital (Roberts, 1974). Aaron (2019) referred to upward communication as the "voice" in organisational behaviour literature, representing the speech having a constructive purpose. As a part of communication in the workplace, upward communication and particularly upward feedback are substantially attributable to organisational performance in common sense. However, for its pure sensitivity in direction of communication and difference in power held by receiver and sender, upward feedback faces radical challenges. Statistics revealed that more than 50% of employees working in the high-pressure working environment such as hospitals prefer to stay quiet (Aaron, 2019). Otherwise, if any, employees tend to hide the truth in their feedbacks. Furthermore, it was found that the problem of upward feedback tends to rise significantly in large organisations wherein the distance of power between hierarchies is considerably higher.

Nordby (2020) advocated that poor upward communication reflects poor organisational culture as those in higher status assume dominant voice, yet the consequence on organisational performance can be massive. Particularly, incorrect information available in upward feedback may drive organisations to the wrong detection of the problem in human resources management (Smith & Fortunato, 2008). Meanwhile, even though employees have the chance to raise their voice, they remain coping with emotional distress as the organisation is misguided by unfaithful feedbacks. Such potential threats invoke the need for research on the method of eliciting truthfulness in upward feedback, through which the relevant practical solutions can be determined.

This research will shed light on the benefits of adopting the Bayesian Truth Serum (Weaver & Prelec, 2013) in the specific regard of eliciting truthfulness from upward feedback. Though the researches on this subject remain limited, the findings by far supported the benefits of BTS as a potential instrument for accessing private information. For its very nature, the BTS helps practitioners create incentives for truth-telling (Weaver & Prelec, 2013); hence, it is highly appropriate to the

problem of upward feedback which involves a low level of truth. According to Prelec (2004), employing the BTS in surveys may render information collection more effective in testing objective knowledge, through which the quality of subjective data is improved. In other words, the BTS is a potentially viable tool for overcoming the existing problem in collecting upward feedback at a desirable extent of truthfulness. Meanwhile, it is imperative to note that the existing literature remains limited in applying the BTS in social-economic contexts. There has been no research specifically directing the BTS to enhance the quality of upward feedback. Therefore, this thesis is dedicated to examining the extent to which the BTS can benefit from eliciting truthfulness in upward feedback, by the virtue of addressing the human resources problem as stated in the earlier section. This primary objective is broken down into a number of sub-objectives for the practical undertaking of research as follows:

i. To examine the way the BTS can be applied in achieving a higher quality of upward feedback, particularly in eliciting truth;
ii. To examine whether the BTS can significantly improve truthfulness in upward feedback;
iii. To figure out the managerial implications for integrating the BTS in addressing the problems related to upward feedback.

The above research objectives are translated into specific research questions which will act as the guide for the practice of the research process as follows:

Q1: How can the Bayesian Truth Serum be adopted in the design of employee surveys for the acquisition of truthful information?

Q2: Are the benefits of adopting the Bayesian Truth Serum statistically significant in eliciting truthfulness of upward feedback?

Q3: What are the implications for the practice of integrating the Bayesian Truth Serum in enhancing upward feedback quality?

Literature Review

Bayesian Truth Serum

Concept of the Bayesian Truth Serum

Theoretically, the Bayesian Truth Serum (BTS) refers to the scoring technique proposed by Prelec (2004) to improve the truthfulness in questions involving personal information, especially for multiple-choice questions. The BTS is practically adapted in surveys via asking each respondent a dual report including (i) an endorsement of the answer to a multiple-choice question and (ii) a prediction of the sample distribution of the corresponding endorsements. The mechanism of the BTS lies in the assignment of the score for an answer that features a higher actual frequency compared to that in prediction. As the prediction derives from the same population, the BTS penalises the answers which are "surprisingly uncommon" while rewarding those being viewed as "surprisingly common" (Prelec, 2004). Under the assumptions of the Bayesian reasoning, the incentive will promote the provision of the truth, which is the best guess that one bears about the surprisingly common answer. Likewise, the unfaithful answer can be minimised for fearing that it might deviate from the ones of the remainders, provided that the others answer truthfully. This treatment, hence, creates incentives for survey takers to provide the truth which reflects the reality of events or subjects to be concerned, even though the true response or expected aggregate response remains unknown.

The application of the BTS is primarily based on the use of an information score for benchmarking and determining whether the answer is truthful (Prelec, 2004). The function of the information score is presented as:

$$iScore = Log\frac{\bar{x}_k}{\bar{y}_k}$$

of which $\bar{x}_k$ denotes the actual relative frequency of response k whereas $\bar{y}_k$ represents the geometric average of frequencies in prediction of k (Prelec, 2004).

In combination with a prediction score, it is possible to produce the BTS score for person r in set of

$$\mathrm{u}^r = \sum_{k=1}^{m} {x_k}^r Log \frac{\bar{\mathrm{x}}_k}{\bar{\mathrm{y}}_k} + a \sum_{k=1}^{m} \bar{\mathrm{x}}_k Log \frac{{y_k}^r}{\bar{\mathrm{x}}_k}$$

This scoring system turns the survey into a "competitive, zero-sum contest", featuring truth in the Bayesian Nash equilibrium (Weaver & Prelec, 2013). Accordingly, the application of the BTS theoretically stimulates the information score to be maximised (for example, respondents are given the incentive to tell the truth), as long as respondents believe that the answer of others is truthful and that the perfect Bayesian predictions are reflected in data distribution (Weaver & Prelec, 2013).

Assumptions of the Bayesian Truth Serum

It is imperative to note that the BTS resorts well to various assumptions to stay valid, therefore, the account of these assumptions might set the ground conditions to test the theoretical application of the BTS in practices. First of all, the theory is rooted in the assumption that personal preference and predicted distribution are positively aligned. However, since participants in the BTS applied survey updated prediction of distribution upon personal belief, it is likely that those who bear homogeneous preference would hold similar anticipation (Weaver & Prelec, 2013). In turn, this also implies that there is a constant connection between the preference or behaviours of one individual and what he predicts the crowd would behave. This assumption stays in line with a number of psychological phenomena, including the False Consensus Effect, which refers to the overestimation of commonness of one's beliefs (Ross, Greene, & House, 1977). Besides, confirmation bias is also another psychological bias that drives people toward developing commonness of belief. Theoretically, confirmation bias explains that people who hold a pre-set belief tend to seek supportive shreds of evidence from the outer world as a way to neglect the contradicted beliefs and any cognitive dissonance arising from the opposite side of view (Aaron, 2019). An individual, consequently, may find himself selectively exposed to information that amplifies his beliefs, inducing the commonness of personal ideas to be exaggerated to a certain extent. All in all, the presence of these psychological phenomena suggests that the BTS's core assumption is pragmatic.

Besides, as long as the Nash Equilibrium is established, it is assumed that an individual predicts other people to think and behave on a rational basis. Such rationality, hence, represents the genuine reflection of reality and set out the ground for a truthful response. Another condition for the BTS is that data collection is conducted with a sufficiently large sample size which possibly renders the impact of a single response or anticipation on the overall result insignificant. Practically, this assumption adds up the requirement for the methodology of study related to the BTS in practice. This requirement is then translated into a specific design of sample size which will be discussed in the methodology chapter.

Applications of the Bayesian Truth Serum

Along with the increase in the number of studies in the BTS, the application of the BTS has been expanded into different contexts in social studies, given the rich possibility of using multiple-choice surveys for collecting subjective feedback. One of the most common pieces of research in the application of the BTS can be referred to that of Weaver and Prelect (2013), wherein the author examined the employment of the BTS through a reward and penalty mechanism. Participants were induced to provide the truth upon the BTS score claimed in recognition of foils within a questionnaire, where one-third were non-existent. Meanwhile, foil recognition was heavily penalised. It was then found that the BTS-based payment was beneficial as a truth-telling incentive as survey takers became less likely to recognise foils, even though the instructor did not provide an explanation. In another study, the effect of the BTS manipulation was examined through odd-ratio of self-admission rates in scientific misconduct compared to that in the control group (John, Loewenstein & Prelec, 2012). The incentive mechanism in such a study was taken under the form of a donation to a charity of respondent's choice, with dishonest answers being subjected to reduction of donation. Under such application conditions, it was suggested that a truth-telling incentive under the form of the BTS design can positively encourage professional respondents to provide honest feedback to questionable research practices.

Another example was the application of the BTS to validate deterrence in criminology by Loughran, Paternoster and Thomas (2014). In this study, respondents were asked to conduct a self-report of particular behaviours such as driving after drinking or cheat on an exam. The application of the incentivising mechanism of the BTS was through examination of the difference between the two groups: one with financial incentive plainly for participation and one with the BTS financial incentive for individuals who are encouraged to be honest (Loughran, Paternoster & Thomas, 2014). It was concluded that the application of the BTS in the self-report survey helps to fetch a higher likelihood of confession in self-report while lowering the estimates of perceived risks related to such commitment.

In another study, Gryz et.al (2015) examined the integration of the BTS in incentivising confession of pirate practice involving digital products. In the output, it was also confirmed that the BTS application as the reward effectively rendered higher willingness of confession: incentivised "pirates" admit approximately 60% more often than the non-incentivised ones.

Nonetheless, past studies found that the BTS integration could bear certain motivational effects in the behaviours of the researched subjects, especially eliciting truthfulness in the areas where people are more likely to provide untruthful feedback, such as criminology or unethical conduct. Going beyond the domain of social study, Howie, Wang and Tsai (2011) proved that there is a high potential for the BTS to be applied in commercial aspects, notably in new product development. Putting in the context of developing pharmaceutical products, the authors found out that the BTS application improved the prediction accuracy of new product adoption by 36% compared to the non-BTS survey.

Interestingly, not every study applied the incentive-based application of the BTS. Instead, a few studies revealed that just integrating the BTS design in the survey can bring forth favourable outputs in truth-telling. For instance, the study of Miller, Bailey and Kirlik (2014) took into account using the BTS for evaluating the degree of learning in design, thus, examining whether the BTS application can help to address the challenge in capturing and assessing what students have learnt. The author proposed a scoring algorithm which was then implemented in assessing participants' responses to the BTS survey. Unlike the original measure of using incentives, the study revealed that

integrating the BTS assessment method can be effective in stimulating the provision of participants for ideas related to the design problem.

All in all, it could be seen that the applications of the BTS in published studies are rather diverse, with proven evidence of success in eliciting truthful answers in different contexts. However, there remains a shortage of study into whether the BTS could be applicable to the purpose of human management, particularly in employee feedback provision. By tapping into this knowledge gap, it is expected to further expand the usefulness of the BTS application.

Upward Feedback

Upward feedback has been one of the most instrumental communication tools on performance evaluation and employee engagement within organisations. Upward feedback concerns the process in which employees offer assessment on their managers' work in confidentiality, then the input would be compiled and presented to the managers with the aim of boosting productivity and planning strategically on the corporate level, to name a few. In fact, according to Gallup (2020), managers are responsible for more than 70% of the difference in employee engagement ratings across business divisions. Since perceiving feedback as accurate plays a crucial part in the management behaviour change (Gates, 2002), it is vital that subordinates provide truthful responses.

Importance of Upward Feedback

According to Tourish and Robson (2004), the importance of upward communication should be viewed through not only benefits to an organisation but also the cost incurring as the firm fails to utilise the upward communication system. On one hand, there have been numerous research papers concerning upward appraisal, with a number of benefits being identified. Of which, most researchers uphold that proper upward communication contributes to promoting the integrity of organisational management, as two-way communication creates the sense of shared leadership (Moravec, Gyr & Friedman, 1993). Consequently, the synergy in the relationship of manager-employees can be improved to generate better efficiency of business operation. The open-door policies nurturing

upward feedbacks also support knowledge management within the organisation, as information can be obtained more effectively from the bottom up.

As far as human management is concerned, the practice of promoting upward feedback may set out the ground for defusing conflict between managers and employees (van Dierendonck et.al, 2007). Accordingly, rather than just delivering feedback or suggestions for business solutions, upward communication can be referred to as a system of expressing complaints or negative feelings. The room for such emotional expression hence renders business an organic organisation where employees can relieve their inner emotions (Tourish & Robson, 2004). Likewise, it is also possible for organisations to gain an understanding of employee's feelings as the reference for the corresponding HR management practices. In specific, there is a wide spectrum of human resources management related to upward communication, including conflict settlement and motivation. According to Stoneman, Bancroft and Hailing (1995), an important benefit of a decent upward communication system is that it opens space for lower-level staff to accomplish high-level needs in Maslow's hierarchy of needs such as self-actualisation. Since employees are given the right to raise their voices and to contribute to organisational performance on a proactive basis, they would eventually feel their presence in the organisation more valuable (Gates, 2002). Consequently, this provides the fundamental for retaining employees in a cohesive relation. The motivational benefits of upward communication also potentially derive from the establishment of a favourable working environment wherein employees are encouraged to develop creativity and innovative thinking. This is highly beneficial to organisations in the rapidly changing contexts of the contemporary business environment which demands ever greater innovation to stay adaptive. Last but probably not least, upward feedback is undeniably the important benchmark for evaluating a manager's performance. This creates certain pressure for managers to try harder and to spend more effort in reinforcing relations with his subordinates.

On the other hand, it is argued by Tourish and Robson (2004) that organisations which fail to include comments of employees may entitle to tremendous costs. The greatest and most visible challenge is low quality of managerial decisions at the top. According to Tourish, D., & Robson, P.

(2004), even in organisations where decision making is centralised, the decision drawn by top managers must be built up from ground information which is only extracted from employees. Therefore, it is concluded by Roberts (1974) that only a strong system of upward communication, including upward feedback to the decisions of managers, can assure effective decision making.

Current Practices of Upward Feedback

Upward feedback is one of the various feedback categories at the workplace, apart from peer feedback or downward feedback. However, while sharing the common feature of information provision for performance appraisal or suggestions for improvement, upward feedback bears a radical disparity from the other types of feedback due to the imbalance in power of feedback provider and feedback receiver. For its high sensitivity, the method conducting upward feedback is also more limited than the other modes of feedback. According to Fleanor and Prince (1997), while downward feedback is less sensitive, there is a multitude of ways for carrying out downward feedback, from annual performance appraisal to direct communication with supervisor. On the contrary, upward feedback is less likely available in these modes of open communication, even though companies motivate employees to freely provide their feedback. Instead, most upward feedbacks are done in the indirect fashion of anonymous feedback (Fleanor & Prince, 1997).

The common approach for organising upward feedback is through an employee engagement survey wherein employees can provide feedbacks toward the manager or company as a whole while keeping their identity confidential. Another approach is through an employee suggestion box where employees may leave anonymous feedback without a digital footprint. However, in the world of innovation, this method of collecting employee feedback becomes far less common, as employees might feel uncomfortable with physical format (Chow, Hwang, & Liao, 2000). Alternatively, there has been greater concern about sending feedbacks through review sites. Accordingly, Glassdoor, InHerSight or CareerBliss are the most popular review sites for employees to post their feedback to employers. As these sites are usually considered as a public reference for evaluating a company's HR management, it could render a double edge for a firm to rely on these sites for freely collecting

feedback. Regardless of which approach is undertaken, it is imperative to see that upward feedback is often available on an anonymous basis. According to Aaron (2019), detaching identity enables employees to feel less burden that feedback provision might cause. However, Gryz et.al (2015) added that confidentiality of personal identity is only the basic condition in truthfulness attainment, as many other cognitive biases may still overtake and drive to less reliable answers even though the identity of employees is unknown. The types of these biases and their impacts on truthfulness in upward feedback will be discussed in the section below.

Behavioural Biases Affecting Truthfulness in Upward Feedback

Although the importance of truthfulness in providing upward feedback is evident, there are factors which act as a hindrance from an employee's perspective. The following section will present four potential reasons affecting the accuracy of the upward feedback at the expense of the company, including relationship.

Manager-employee relationship. According to Bol (2011), feedback represents the subjective idea of one person, hence, it is impossible to separate subjective feelings of an individual from feedback provision. However, the truth represents the objective presence of facts in the real world. The difference between subjective sentiment and objective fact renders untruthful feedbacks to be a misleading representation of reality, henceforth the function of feedback in fact reporting become ineffective. Bol (2011) found out that untruthful feedback may occur with both positive and negative feelings toward the manager, yet the former case appears to be far more common than the latter. Accordingly, the bias of positive feelings toward the manager would result in higher evaluation (Tsui & Barry, 1986; Robbins & DeNisi, 1994), though it cannot completely exclude cases where a bad relationship with managers induce employees to report negative feedback despite the good performance of managers.

Retaliation. It is suggested that employees may fear that providing honest feedback to their managers could cause retaliation subsequently, and thereby have a negative impact on their career advancement (Barrow, Agius & Baker, 2013). According to Aaron (2019), in the process of

contemplating dissent, employees often consider whether their ideas would be viewed as a constructive or destructive contribution to the manager or organisation. The latter case generally involves retaliation from managers, regardless of direct or indirect delivery of retaliation. Should employees perceive their feedback as irrigation to the manager and possibly invokes retaliation against them, there is a tendency of hiding the painful truth to avoid the foreseeable bad consequence. Instead, employees tend to opt for easily acceptable feedback which favours their status quo, as they choose risk aversion. The study of Smith and Fortunato (2008) categorises the fear of retaliation as a mediator for the intention of providing upward feedback. It was found out that people who perceive their feedback as the potential cause of retaliation risks would be more likely to provide dishonest trust due to other direct factors such as cynicism.

Cynicism. Theoretically, cynicism is referred to as a "negative attitude toward one's employing organisation" (Smith & Fortunato, 2008). The term is commonly adopted to reflect the state that employees are cynical about the implication that feedback would bring forth. As the extent of cynicism increases, it is less likely that employees trust the impact that their feedbacks associate with. In other words, should the employees expect or understand that little to no change would be made after the feedback, they might be reluctant to respond to the evaluation truthfully (Smith & Fortunato, 2008). Statistically, 90% of employees reported that they are more likely to stay at a company which "takes and acts on feedback" (Achievers, 2020). This somehow explains why organisations with no transparent and relevant response to the output of feedback would take less employee enthusiasm for an internal survey; hence the quality of upward appraisal might deviate significantly. Instead, businesses that feature responsive follow-up plans enjoy the higher effort of employees in delivering truthful feedback, as one might fear that any dishonest response would harm the subsequent managerial decisions.

Leniency. Leniency bias refers to the tendency of giving an easy rating on other people for avoiding negative impacts on the people being appraised (Bol, 2011). In a business context, leniency might take place to both upward or downward feedback, though the latter case appears to be more

common in practice. However, considering the context of upward feedback, leniency bias often arises from the avoidance of confrontation with the managers and disclosure of the current problems related to a negative rating. According to Chow, Hwang and Liao (2000), positive feedbacks are more easily accepted by the organisation without further tracking for the root cause of problems, which happen with negative feedbacks. Not to mention that positive feedback appears to reflect the more positive performance of the team as a whole; thus, employees benefit more from providing good ratings even though this is misleading feedback or not what genuinely occurs.

Bridging the Literature Gap

While there are different biases potentially contributing to the hindrance of truth provision in upward communication, the BTS is potentially applicable to give truth-telling incentives even though it does not directly mitigate these biases. Because the BTS adopts the assumption that the overall result is not affected by a single prediction or response, the biases related to the relationship between managers and employees or leniency can be somehow addressed. However, certain biases might need a specific setting of upward communication to take effect. For instance, the fear of retaliation can only be avoided in an anonymous survey while cynicism often depends more upon the managerial changes associated with the collection of upward communication. That being said, the use of the BTS merely contributes to increasing the likeliness of truth provision, rather than being an all-in-one solution which can rule out the unfaithful answer. However, there existed an absence of research examining this effect by far, as the application of the BTS remains limited in incentivising truthful responses to the generic subjective survey (Frank, et.al 2017) or collection of information from the social study. The implementation in the business context is rather scarce with the main application lies in a consumer survey for preference (Radas, 2015) whereas no research connecting the BTS and upward feedback. Considering this gap in the literature and the potential of using the BTS for overcoming the current problems of upward feedback, especially the challenges of eliciting truthful responses from employees. Even when it was revealed that truth provision in upward feedback is hindered by a number of factors as mentioned above, the limited availability of research also makes it worth

investigating whether the BTS can address the problem from its root by tackling these identified biases. Such confrontation to the unavailability of knowledge from the existed literature renders this research more valuable.

Experimental Design

Experimental Strategy

In line with the primary objective of examining the benefits of the BTS in generating truth incentives for upward feedback, the experimental study design will be employed. This research design features a number of advantages including a stronger hold over variables or high efficiency in testing the relation between variables in a controlled environment (Chow, Hwang & Liao, 2000). Most importantly, the experimental design enables the study comparable to the existing findings from the vast majority of past research papers conducted in this approach.

Practically, the experiment will be conducted with the participation of research subjects who are employees of large organisations with high employee turnover rates. It is assumed that a large organisation extends the distance of power and, hence, renders a tougher obstacle to upward communication. Research participants will be then asked questions of their feedback about different topics, ranging from relation with the direct supervisor to the comment about the manager's performance in HR management. The participants will be separated into two groups to test the effectiveness of the BTS integration in the employee survey. Accordingly, the control group will receive a normal questionnaire of the non-BTS application, whereas the experimental group will take the BTS integrated questionnaire. For motivating the engagement of respondents in data collection, both treatment and control groups will be offered monetary rewards under different conditions. Participants in the former group who scored the highest BTS index will also have a chance to receive the reward of €10. For the latter group, the reward of €10 will be drawn on a random basis of the random lottery incentive.

Data Collection

Questionnaire Structure

The study will employ a structured questionnaire to achieve a higher consistency of data set whilst the delivery of the questionnaire is also streamlined and simplified. The use of a structured questionnaire also facilitates the process of data cleaning and extraction for pattern analysis, through which the researcher can quickly draw an understanding of data and establish knowledge from data analysis. Practically, with the research design as proposed above, there will be two sets of questionnaires for the BTS and non-BTS group, with the only difference between the two sets is the addition of prediction questions exclusively available in the BTS questionnaire. The core structure of each questionnaire, regardless of the BTS or non-BTS version, will include three sections: (1) demographic information of respondents, (2) upward feedback to respondent's manager and (3) investigation of biases in upward feedback. Of which the first section will comprise various multiple-choice questions to get information on respondents taking part in the data collection session. The second section encompasses various yes/no questions to collect data for truth testing of the BTS application. The last section includes questions regarding the response of employees to biases hindering truthfulness in upward feedback. To capture the response of employees, the 5-level Likert scale will be deployed (Appendix 1).

Data Collection Strategy

The research primarily examines the effectiveness of applying the BTS in eliciting the truthfulness of upward feedback. Due to the unavailability of secondary data from similar papers, it is necessary to collect primary data for answering the research questions proposed at the beginning. Particularly, for the nature of the BTS as an instrument compatible with multiple-choice surveys, the research will examine data collected from a survey. In practice, the survey will be separated into non-BTS and BTS surveys to draw conclusions upon the comparison of the two groups.

While there is a multitude of routes for survey data collection, it is ideal to employ an online survey to avoid the geographic constraint, which may happen in a face-to-face survey. This instrument is also relevant since other alternatives such as telesurvey or email surveys tend to require a good database. Google Form or Qualtrics will be used to compose the survey questionnaire for its

advantages of easy distribution and extraction. The survey will be distributed on a professional social network which attracts people having an interest in human resources and organisational behaviour problems, such as LinkedIn. For backup, the post of the survey link on other social networks such as Facebook or Twitter is also in consideration.

Sampling Strategy

The target subjects will be employees in organisations operating in the Netherlands where the researcher is residing and studying. The scope of the survey is limited in the Netherlands for the ease of data collection, while it can remove the bias or distortion in feedback due to differences, which might happen in a cross-culture context. Other than these conditions, there will be no criterion for selecting research respondents.

The research uses particularly purposive sampling technique, a non-probability sampling approach having strength lied in convenient access to sample (Scheurich, 2014). This makes it feasible to save effort in data collection (Wellington & Szczerbinski, 2007). Also, this technique eases the set of samples, considering a focus on certain niche demographics. However, a low response rate could be referred to as the disadvantage of purposive sampling (Cohen, Manion & Morrison, 2011). To cover this drawback, snowball sampling might be used as a backup. Snowball sampling implies participants would introduce others who they perceive or know as having potential interest in the research topic (Sharma, 2005).

Selection of sample in research is crucial to the quality of research outcomes. There are a number of approaches to the determination of sample size. Theoretically, the sample size should be at a minimal line of 30, for bypassing the assumption of normal sample distribution under the Limit Central Theorem (Scheurich, 2014). The BTS assumes a large sample size by default as no single response may influence the overall results. For meeting this condition, this research initially aimed to adopt a sample size of 100, equally split between the control group and the treatment group. In the actual data collection session, it was possible to collect 211 observations. While this sample size was sufficient to assure the normal distribution for each group in line with the condition of Limit Central

Theorem but was also sufficiently large to assure representativeness of the sample in population. Besides, this sample size is too large to make it neither impractical nor unachievable. To test the adequacy of this sample size in producing a clinically meaningful prediction, the power analysis was taken (Appendix 3). Accordingly, it was found that the desired power can be achieved with the minimal sample of 162 with 81 samples in each group. This implies that the sample size of 211 as per the actual collection of this study should be qualified for producing statistical meaning.

Hypotheses

These experimental settings will help test whether the use of the BTS modifies the tendency of answers in upward feedback. Also, it is possible to test the specific benefit of the BTS application in tackling the hindrance of truth-telling, under the lens of bias motives. The hypotheses for the study will be presented hereunder.

The hypothesis for the impacts of the BTS on upward feedback is presented below.

H1: The application of the BTS in employee survey statistically significantly decreases positive response in upward feedback

This hypothesis is produced on the ground assumption that the frequency of negative response represents the extent of higher truthfulness, as the employees dare to speak up about matters which might involve some certain level of risk.

The hypotheses for benefits of the BTS in mitigating psychological bias affecting truthfulness of upward feedback are presented below.

H2: The application of the BTS in employee survey statistically significantly lowers retaliation biases in manager-employee relationship in upward feedback provision

H3: The application of the BTS in employee survey statistically significantly lowers relation biases in upward feedback provision

H4: The application of the BTS in employee survey statistically significantly lowers cynicism biases in upward feedback provision

H5: The application of the BTS in employee survey statistically significantly lowers leniency in upward feedback provision

Ethic Consideration

The inclusion of survey as a data collection method renders this study human involvement research. This requires the research to abide by the standard requirements for research ethics in line with the Erasmus School of Economics's Code of Practice Governing the Ethical Conduct (Appendix 2).

Data Analysis

Data Description

The study encompasses 211 independent observations with 103 (or equivalently 48.82%) participants subjected to the BTS treatment; whilst there are 108 (51.18%) participants in the control group to reflect the effect of the non-BTS treatment. Table 1 presents the summary of variables and data adopted in the study.

Table 1

Data Description

Variable	Question	Type	Description	Range
BTS		binary	Indicates if the participant answered the questionnaire under the BTS condition (1) or not (0)	0-1
Year_of_work	Q4.1; Q4.2	categorial	Indicates number of years working 1: Less than 1 year; 2: 1-3 years; 3: 3-5 years; 4: more than 5 years	1-4
Size_of_business	Q5.1; Q5.2	categorial	Indicates size of business 1: Small; 2: Medium; 3: Large	1-3
Constructive_Feedback Autonomic_Decision Personal_Care Management_Expertise Goal_Communication Accountable_Manager Task_Completion	Q9 Q10 Q11 Q12 Q13 Q14 Q15	binary	Indicates if the participants answered the questionnaire Yes (1) or No (0)	0-1
Predict_Constructive_Feedback Predict_Autonomic_Decision	Q9.1 Q10.1	continuous	Indicates the prediction of participants for the answer	0-100

Predict_Personal_Care Predict_Management_Expertise Predict_Goal_Communication Predict_Accountable_Manager Predict_Task_Completion	Q11.1 Q12.1 Q13.1 Q14.1 Q15.1		to be given by other people for questions 9-15	
Subjective Relation Retaliation Cynicism Leniency	Q16 Q17 Q18 Q19	categorical	Indicates the level of agreement 1: Strongly disagree; 2: Somewhat disagree; 3: Neither agree nor disagree; 4: Somewhat agree; 5: Strongly agree	1-5
Positive Feedback	Average of Q9-Q15	continuous	Measure upward feedback. It is assumed that a higher value represents more positivity of feedback	

Descriptive Statistics

Descriptive Statistics for Demographic Variables

The first group of data includes demographic variables (year of work and size of business). Table 2 illustrates that the distribution of data for these variables bears an insignificant difference between the BTS and non-BTS group as the t-test two-tailed P-values for the difference are captured at 0.45 and 0.56, respectively. The high level of homogeneity renders a better condition for testing the BTS as it can assure that any differences in upward feedback (if relevant) arising from the two groups will derive from the effect of treatment rather than demographic differences.

Table 2

Descriptive Statistics of Demographics

	Non-BTS		BTS		
Variable	*Mean*	*SD.*	*Mean*	*SD.*	*Two-tailed P-value*
Year of work	1.67	1.45	1.53	1.48	0.45
Size of business	2.43	0.79	2.23	0.83	0.56

Descriptive Statistics for Upward Feedback

Table 3 shows the comparison of the probability of responding “Yes” to different items in upward feedbacks between the treatment and control groups using t-test. It reveals a significant difference between the two groups, as the average in the control group is generally lower. In other

words, there exists a tendency of providing more negative upward feedback in the treatment group compared to the control group. To evaluate the effect of BTS manipulation in a statistical manner, the study adopted an odd ratio as what John, Loewenstein and Prelec (2012) did. Theoretically, the odd ratio represents the measure of association for a case-control study, calculated upon the quantity of exposure to a factor and those in the control group (Pearce, 1993). The odd ratio also reflects this substantial difference, while the P-values for the difference of all items are all below 0.05. However, these results are somewhat different from those in the study of John, Loewenstein and Prelec (2012), wherein only one item captured a high odd ratio. Certainly, the difference in subject of study could be attributable to this difference. It can be roughly concluded that the BTS manipulation effect for upward feedback is high for all aspects.

Table 3

Upward Feedback Summary Between the Treatment and Control Groups

	Percentage of "Yes"		Odds	Odd ratio	Two-tailed P
	BTS	*Non-BTS*			
Constructive_Feedback	18.4%	42.1%	1.395	3.168	0.000454
Autonomic_Decision	20.3%	50.5%	1.000	3.682	5.54E-06
Personal_Care	26.2%	46.3%	1.160	2.463	0.00336
Management_Expertise	30.1%	51.2%	0.964	2.465	0.002361
Goal_Communication	29.1%	45.1%	1.204	2.063	0.02091
Accountable_Manager	26.21%	42.0%	1.4	1.946	0.038194
Task_Completion	18.44%	51.5%	0.964	4.299	1.09E-06

To examine the impact of the BTS from a more systematic view, the study employs a z-test for determining whether the differences in the mean of respondents' responses statistically vary between the treatment and control group. The test is in response to the hypothesis:

$$H_0: \bar{x}_1 - \bar{x}_0 = 0$$

$$H1: \bar{x}_1 - \bar{x}_0 \neq 0$$

Refer to table 3, the P-values captured for all questions are below 0.05, hence, indicating that the application of the BTS statistically induces lower positivity in upward feedback to the managers. Consequently, this rejects the null hypothesis of "*H1: The application of the BTS in employee survey statistically significantly decreases positive response in upward feedback"*, which implies the potential

benefit of adopting the BTS mechanism in eliciting truthfulness out of upward feedback survey. To further investigate the nature of influence, it is imperative to consider the interaction between the BTS application and the likelihood of subjective biases in upward feedback.

Descriptive Statistics for Biases

Table 4 displays that on average, the BTS condition group score higher on the subjective relation bias by 0.15 points relative to the non-BTS group, but both groups exhibit a similar standard deviation. The mean and standard deviation for the retaliation and leniency are very similar for both groups. In contrast, the BTS condition group score lower on the cynicism bias by 0.03 points, relative to the non-BTS group, but both groups also exhibit a similar standard deviation. The t-test two-tailed P-value reflects that only the difference in committing to subjective biases see a statistically significant difference between the treatment and control group at 95% confidence. The remaining categories, however, capture an insignificant difference in the average commitment between the two groups.

Table 4

Descriptive Statistics of Interval Variables between the Treatment and Control Groups

Variable		Non-BTS		BTS		Two-tailed P
		Mean	*SD*	*Mean*	*SD*	
Subjective Relation	*Q16*	2.86	1.19	3.01	1.36	0.04
Retaliation	*Q17*	2.64	1.21	2.69	1.35	0.26
Cynicism	*Q18*	2.68	1.23	2.65	1.24	0.45
Leniency	*Q19*	2.50	1.14	2.47	1.17	0.31

Correlation Analysis

Table 5

Correlation of Biases and Upward Feedback in the Control Group

	Positive Feedback	Retaliation	Relation	Cynicism	Leniency
Positive Feedback	1.000				
Retaliation	0.559*	1.000			
P-value	0.000				
Subjective Relation	-0.202*	-0.106	1.000		
P-value	0.037	0.276			
Cynicism	0.758*	0.404*	-0.132	1.000	
P-value	0.000	0.000	0.172		
Leniency	-0.066	-0.149	0.044*	-0.162	1.000
P-value	0.497	0.125	0.000	0.093	

*Note: * p<0.05*

In absence of the BTS treatment, it appears that most of the biases show an average to strong correlation with upward feedback (Table 5). Of which, the correlation coefficients of cynicism and retaliation with upward feedback were captured at 0.556 and 0.758, respectively. This suggests that respondents in the non-BTS group tended to provide more positive upward feedback when experiencing these types of biases. However, the correlation coefficients of relation bias and leniency were found at the negative sign and far lower magnitude.

Table 6

Correlation of Biases and Upward Feedback in the Treatment Group

	Positive Feedback	Retaliation	Relation	Cynicism	Leniency
Positive Feedback	1.000				
Retaliation	0.395*	1.000			
P-value	0.000				
Subjective Relation	-0.046	0.0176	1.000		
P-value	0.647	0.860			
Cynicism	-0.173	-0.083	-0.003	1.000	
P-value	0.081	0.407	0.974		
Leniency	0.014	0.060	0.063	0.050	1.000
P-value	0.887	0.522	0.503	0.616	

*Note: * p<0.05*

In the treatment group, the correlation coefficients between the biases and upward feedback record a comparatively lower magnitude of relation (Table 6). In particular, only retaliation is found to have a correlation coefficient of approximately 0.4, whereas that for cynicism bias is only 0.17. Relation and leniency have correlation coefficients with upward feedback of less than 0.1, thus, indicating very weak connections. For the sign of relation, retaliation and leniency are found to bear a positive correlation, unlike negative coefficients among the remainders.

Another important feature is that the correlation coefficients among the biases are quite weak, mostly under 0.1 in both the control and treatment groups. Therefore, it can conclude that multi-linearity is not available in the dataset, which renders it possible to run a regression analysis.

Results

Prevalence Estimates

To test prevalence estimates, the study of John, Loewenstein and Prelec (2012) adopted geometric means of variables as “Conservative judgements of true prevalence”. The geometric means of estimate for the "Yes" answer, given that the person provided “Yes” or “No” for the corresponding items were provided as in Table 7. Respondents who answered “Yes” would constantly opt for the prediction of seeing a similar answer in other people. This can be visibly seen through the fact that the geometric mean of “Yes” from the respondent who answered Yes reached more or less 80% for all items, twice as many as that of the counterpart. The figure for prevalence estimate of “Yes” from respondents who answered “No” was as low as less than 30% in some specific items such as management expertise or task completion. On this stark contrast, the assumption of BTS likely holds. This denotes that individual who opts for a certain response will be likely to provide a higher prediction for selection of the others for the same response, compared to the likelihood of selecting the remaining alternatives.

Table 7

Results of Testing the BTS Assumption for Each Question

	Geometric mean of prediction for “Yes” from respondent who answered “Yes”	Geometric mean of prediction for “Yes” from respondent who answered “No”
Predict_Constructive_Feedback	86.46	41.76
Predict_Autonomic_Decision	83.18	41.95
Predict_Personal_Care	83.59	42.13
Predict_Management_Expertise	81.03	29.51
Predict_Goal_Communication	84.64	43.02
Predict_Accountable_Manager	79.40	31.37
Predict_Task_Completion	80.96	30.81

Impacts of the BTS Application in Moderating Relationship between Cognitive Biases and Truthfulness of Upward Feedback

To test whether there exists a relationship between cognitive biases and upward feedback under both conditions, a regression test is performed with the independent variables being fixed to

the four types of biases identified in the literature review. The logic behind this test is to identify the differences in upward feedback upon the changes of biases between the BTS and non-BTS conditions.

For the control group

The R-squared is captured at 0.675, showing that involvement to biases explains up to 67.5% change in upward feedback in the control group. This represents a relatively high level of relevance for studying upward feedback accuracy from the perspective of cognitive bias, which is confirmed by a high F-value while the P-value of the model approached zero.

Regarding the coefficients, it is established that all variables have a P-value of less than 0.05. In other words, the availability of all biases contributes to the changes of upward feedback. Notably, since the coefficients of all variables are positive, it appears that respondents in the control group would give positive upward feedback as long as experiencing the given biases.

Positive Feedback	Coef.	SE	t	P>|t|
Retaliation	0.0579205	0.0115904	5.00	0.000
Subjective Relation	0.0343228	.0142494	2.41	0.018
Cynicism	.1387768	.0134596	10.31	0.000
Leniency	.0362458	.0152478	2.38	0.019
cons	-.0195349	.0752329	-0.26	0.796

For the treatment group

The R-squared is captured at 0.18, indicating that involvement to biases merely explains up to 18.01% change in upward feedback in the treatment group. This represents a relatively low level of impact that cognitive biases generally have with upward feedbacks.

Positive Feedback	Coef.	Std. Err.	t	P>|t|
Retaliation	.0609902	.0147156	4.14	0.000
Subjective Relation	-.0139774	.0192351	-0.73	0.469
Cynicism	-.0482382	.0308501	-1.56	0.121
Leniency	.0097976	.0222752	0.44	0.661
cons	.7228795	.1248783	5.79	0.000

In terms of the specific regression of each bias, all variables, except retaliation, bear a statistically insignificant relationship with upward feedback at the 5% significance level. Therefore, with the application of the BTS, the impacts of relation, cynicism and leniency on the likelihood of providing positive upward feedback is no longer important. This leads to the acceptance of hypotheses

related to these variables, namely H3, H4 and H5, which propose that the application of the BTS in employee surveys would statistically significantly lower the subjective relation, cynicism biases and leniency in upward feedback provision, respectively.

Only retaliation has a P-value approaching zero, indicating that retaliation still has a significant impact on upward feedback among participants of the treatment group. This leads to the rejection of the hypothesis *"H2: The application of the BTS in employee survey statistically significantly mitigate impacts of retaliation biases in manager-employee relationship in upward feedback provision*."

Comparatively, the connection between biases and upward feedback demonstrates significant changes between the treatment and control groups. While cognitive biases mostly hold a positive influence on upward feedback in the control group, people who experienced the BTS incentives did not have their upward feedback affected by cognitive biases, except retaliation. It can be concluded that the BTS incentive shows an effective impact on mitigating most of the cognitive biases, except the retaliation one.

Discussion

In the face of challenges from subjective biases, which appear to hinder the practice of upward feedbacks, the study was taken to investigate the application of the BTS design for employee surveys with a view to achieving a higher level of truthfulness. The experimental study was conducted through 211 observations to determine the effect of the BTS integration in practice. The results from the data analysis will be discussed thoroughly in the below section.

Claim 1: The application of the BTS in employee surveys statistically significantly decreases the likelihood of positive responses in upward feedback

The experiment with a fair split between the control and treatment group of more than 100 observations in each group revealed a significant difference in the provision of upward feedback across different groups. Accordingly, the treatment group remarked significantly lower positive upward feedback when respondents were assigned to undertake the survey with a BTS-based incentive mechanism. A lower rate positivity, although does not necessarily imply a higher level of

accuracy of the answer, maybe assumed to reflect that employees are more open to provide negative feedback toward their direct managers. According to Smith and Fortunato (2008), employees are more willing to provide positive comments even though such responses do not represent the truth as the positivity is easier to be accepted by society. The application of the BTS in incentivising respondents to give more negative feedbacks, therefore, can be acknowledged as the practical motivation of truth provision. This is in line with the argument of Miller, Bailey and Kirlik (2014) that the BTS application can encourage respondents to overcome the constraint of honest answers in sensitive conditions. This finding, though have yet proven the benefit of BTS application in upward feedback quality improvement, may serve as the important foundation for further investigation into the nature of the BTS for employee surveys in general.

Claim 2: The application of BTS can statistically reduce the involvement of cognitive biases

This claim was examined through four hypotheses which encompass the four types of biases that potentially hinder the truth provision in feedback toward managers. As identified in the literature review, these biases include subjective biases in the manager-employee relationship, retaliation, cynicism and leniency. By examining the data with a statistical comparison between the BTS group and the non-treatment group, it was revealed that the BTS does not hold a significant correlation with every identified type of bias. Particularly, it was found that the application of the BTS incentive appears to bear the most noticeable impact in mitigating certain types of bias, including manager-employee relationship, cynicism bias and leniency. This conclusion was retrieved from the result of a significantly low level of these biases in the BTS treatment group compared to those of the control group. On the other hand, retaliation is the only bias which would not be influenced by the application of the BTS in practice.

From the theoretical perspective, retaliation bias differs significantly from the remainders as it involves the fear of facing risks for the provision of truthful answers. Therefore, while the incentive mechanism of the BTS can promote truthfulness by rewarding the truth, it appears that the real benefit lies in the compensation for making truthful answers. According to Miller, Bailey and Kirlik

(2014), the incentive mechanism of the BTS derives from both financial inventive for the survey respondents plus the psychological motivation from the crowd effect (as the participants assume that others will also provide the truthful response). Such a complex mechanism can directly enable the survey participants to overcome psychological barriers, which generally induce the provision of a lie or dishonest response. For certain cognitive biases such as relationship with managers or the sense of harming others (leniency bias), this mechanism may work effectively as respondents feel less pressured for relational impacts arising from their disclosure of unfavourable information (Chow, et.al, 2000). In other words, the barrier to truthfulness is delicate enough for the BTS incentive to break through. On the other hand, according to Smith and Fortunato (2008), the fear of retaliation (retaliation bias) is not addressed completely different from the other biases as it involves the probable threats to the information provider. Therefore, unless there is a sufficiently large value in a trade-off for the risks to be taken, it is unlikely that the BTS designed survey can mitigate such type of bias in practice (Smith & Fortunato, 2008). However, due to the financial constraint of this study and the limited scope, it was unable to test whether a higher financial incentive amount would work. This will be adopted as a research limitation for the future study.

Practical Implications

One major motive for the study to be undertaken is to address the challenge of collecting truthful answers for upward feedback, which currently bears difficulty for many organisations. As being shown in the literature, there has been an increase of concern on using review sites for the ineffective use of employee engagement surveys. However, because these review sites propose certain limitations in implementation as a regular internal practice, it would be rather effective for firms to resort to employee engagement surveys. Yet, overcoming the barriers of truth provision and particularly the cognitive biases leading to dishonest answers in upward feedback is the most urgent task.

With the findings drawn out from this study, it is imperative to acknowledge the importance of applying the BTS in mitigating the factors causing dishonest answers. Following the finding that the

BTS incentive mechanism can trigger the willingness of providing more negative comments on managers, there is a potential of adopting this approach in managerial practice, especially in figuring out current problems in management and organisational deficiency. Nonetheless, when it comes to the practical implementation of the BTS in the employee survey, it is also worth noting that the BTS cannot fully address every problem related to cognitive bias as only certain biases are relevant to the effect of the BTS. Therefore, the BTS should not be an all-in-one solution for improving the truthfulness of employee surveys. Instead, the design should only be considered as a standard for constructing an employee engagement survey, followed by relevant addition of other measures of truth-motivation in data collection. In particular, because retaliation bias remains outside the effective domain of the BTS incentive, organisations would need to practice additional solutions which potentially assure the protection of honest information providers. What is more, to take the desirable benefits from the BTS, organisations also need to consider the size of incentive to be rewarded for truth provision. According to Frank et.al (2017), the reward should be sufficiently large to cover any potential risks that employees may be exposed to for making unfavourably truthful answers. Another important point to remark is the design of the data collection session which can directly associate with the confidence of the survey participants. Since upward feedback generally involves sensitive information, organisations need to take sufficient and relevant measures in protecting the anonymity of the employees. This should help to reduce the sense of fear for retaliation, in the event that respondents wish to report negative truth about their manager.

Furthermore, the fact that the BTS can effectively enable survey-takers to overcome specific biases such as the relationship between manager and ethical sense arising from disclosure of unfavourable information also reflects the radical benefits of the BTS design. Technically, the BTS, therefore, can be integrated into an employee engagement survey for accessing a generally higher level of truthfulness. Yet it is also worth noting that the transparency in truth score calculation and rewarding mechanism would be crucial factors assuring the successful undertaking of the BTS. Hence,

organisations applying this technique should clearly concern about these technical terms, before communicating well these points to the survey participants.

Limitations and Future Study Suggestions

Though the study has achieved the objectives which were predetermined, there remains room for improvement due to the presence of certain study limitations. First of all, a caveat of this study derives from the use of a quantitative approach in measuring the likelihood of the BTS effect in truth eliciting. On the one hand, this approach enables the study outputs to stay aligned with the past studies in this area, paving way for comparison and output validation. On the other hand, a quantitative study design bears a conventional weakness in accessing in-depth understanding. Therefore, though the study proved that the BTS application can be relevant to upward feedback in general, it may not as effective in a particular case study with certain settings of organisational structure or relationship between employees and managers. For future study, it is worth investigating the application of the BTS in a specific case study, through the adoption of a qualitative approach to uncover more about this issue. Another limitation of the study is that it lies on the assumption of truthful answers. Since it is impossible to know the honest answer beforehand, which is also considered as an assumption of the BTS, it was then assumed that a more truthful answer associates with a higher level of negativity in the employees' responses. This derives from the fact that it is more challenging and pressurised for employees to deliver unfavourable comments about her/his manager. However, certainly, this is not always the case although the likelihood of putting forth a positive comment is generally more common than negative ones. In other words, the findings and results of this study only stand firmly and validly as long as this assumption is held. The consistent reliance on the assumption, hence, implies that there existed a chance for error in predicting the effect of the BTS application in practice. Last but not least, though it was concluded that the BTS is beneficial to upward feedback, the application incorporate practice remains ambiguous. Nonetheless, there remain questions to be answered before the application of BTS incentives in the employee survey can be initialised. For instance, it is wondered how much the incentive pronounced by the BTS should be for

effectively mitigating each type of biases. This sets forth the orientation for further investigation between the amount of financial incentive and the quality of upward feedback.

Conclusion

The study was undertaken with the primary purpose of investigating the application of the BTS in eliciting truth for improving the quality of upward feedback. The study was meant to address the difficulties in collecting feedbacks toward managers, which appear to be hindered by various factors. The in-depth research of the related literature revealed that a great challenge to truth provision in upward feedback refers to a number of cognitive biases, ranging from relational bias, leniency bias, cynicism bias and retaliation bias. Given the common occurrence of these biases, there have been various techniques being applied in upward feedbacks, yet none has shown absolute convenience and efficiency. Therefore, on the basis of the BTS's benefits in eliciting truth, the experimental study was conducted for examining the relevance of applying the BTS design in the employee survey. Data collected from a total of 211 observations, with 103 in the treatment group and 108 in the control group showed that the implementation of the BTS can significantly motivate the willingness of employees in providing negative comments about managers. This indicates the positivity in the application of the BTS for upward feedback collection. To understand the mechanism of the impact that the BTS design features, the correlation and regression analysis were conducted to understand how the relationship between biases and truthfulness in upward feedback has been moderated. Accordingly, it was found that the application of the BTS helps significantly reduce the availability of impacts that cognitive biases caused in the process of providing upward feedback. However, a crucial finding was that not every bias can be mitigated by applying the BTS incentive mechanism, as retaliation bias remains unaffected. In other words, while the BTS can be effective in encouraging truthful responses for upward feedback in most cases, it unnecessarily guarantees a certain provision of truthfulness, depending upon the perception of retaliation that survey takers feature. If employees perceive higher threats of retaliation for providing unfavourable feedback or the sense of being retaliated significantly affects the likelihood of truthful answers, it is unlikely that

the BTS incentive can work. Therefore, it is recommended to provide additional measures supporting the mental confidence of survey takers for the assurance of a higher level of truthfulness.

References

Aaron, G. (2019). *Challenges and enablers of upward communication: How to foster a speak-up culture in your organization.* Retrieved from: https://www.ckju.net/en/dossier/challenges-and-enablers-upward-communication-how-foster-speak-culture-your-organization/1256

Abolmagd, M. (2017). Tell Me the Unverifiable Truth – Bayesian Truth Serum and Social Desirability Bias: An Experimental Analysis. *Erasmus School of Economics: Master thesis*.

Achievers. (2020). *Engagement and Retention Report: Failure to Engage* [White Paper]. Retrieved from https://www.achievers.com/resources/white-papers/2020-engagement-retention-report/

Barrow, P., Agius, S., & Baker, P. (2013). Trainee perceptions of upward feedback in general surgery. *International Journal Of Surgery*, *11*(8), 689. doi: 10.1016/j.ijsu.2013.06.545

Bol, J. (2011). The Determinants and Performance Effects of Managers' Performance Evaluation Biases. The Accounting Review, 86 (5), 1549-1575

Berk, R., Naumann, P., & Appling, S. (2004). Beyond Student Ratings: Peer Observation of Classroom and Clinical Teaching. *International Journal Of Nursing Education Scholarship*, *1*(1). doi: 10.2202/1548-923x.1024

Chow, C., Hwang, R., & Liao, W. (2000). Motivating Truthful Upward Communication of Private Information: An Experimental Study of Mechanisms From Theory and Practice. *Abacus*, *36*(2), 160-179. doi: 10.1111/1467-6281.00058

Cohen, L., Manion. L., & Morrison, K. (2011). *Research methods in education*. London: Routledge

Couper, M. (2017). Some Methodological Uses of Responses to Open Questions and Other Verbatim Comments in Quantitative Surveys, *Journal of Research, 11(2)*, 112-119

Fleanor, J. & Prince, J. (1997). Using 360-degree feedback in organizations. Center for Creative Leadership

Frank, M., Cebrian, M, Pickard, G. & Rahwan, L. (2017). Validating Bayesian truth serum in large-scale online human experiments, *PLOS One, 12(5),* doi: 10.1371/journal.pone.0177385

Gallup. (2020). *Employee Engagement and Performance: Latest Insights from the World's Largest Study*. Retrieved from https://www.gallup.com/workplace/321032/employee-engagement-meta-analysis-brief.aspx

Gates, K. (2002). *Upward Feedback: Factors Affecting Acceptance and Behaviour Change in CEOs* (PhD). Illinois Institute of Technology.

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

Howie, P. J., Wang, Y. & Tsai, J. (2011). Predicting new product adoption using Bayesian truth serum, *Journal of Medical Marketing, 11(1),* p6-16

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.

Kukla-Gryz, A., Tyrowicz, J., Krawczyk, M. & Siwinski, K. (2015). We All Do It, but Are We Willing to Admit? Incentivizing Digital Pirates' Confessions, *Applied Economics Letters, 22(1-3),* pp. 184-88

Kumar, R. (2010). *Research Methodology: A Step-by-Step Guide for Beginners*. London: Sage publishing

Loughran, T., Paternoster, R. & Thomas, K. (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)*, p677-707

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)*, pp. 487-515

Moravec, M., Gyr, H., and Friedman, L. (1993) 'A 21st century communication tool'. HR Magazine, 77-81

Nordby, H. (2020). Communication and organizational culture. Retrieved from: https://www.intechopen.com/books/a-closer-look-at-organizational-culture-in-action/communication-and-organizational-culture

ODKScheurich, J. (2014). *Research Method in the Postmodern*. Psychology Press

Prelec, D (2004). A Bayesian Truth Serum for Subjective Data. *Science New Series*, *306*, 462-466

Pearce, N. (1993). What Does the Odds Ratio Estimate in a Case-Control Study?. *International Journal of Epidemiology, 22(6)*, pp. 1189-1192. doi: 10.1093/ije/22.6.1189

Robbins, T., & DeNisi, A. (1994). A closer look at interpersonal affect as a distinct influence on cognitive processing in performance evaluations. *Journal Of Applied Psychology*, *79*(3), 341-353. doi: 10.1037/0021-9010.79.3.341

Roberts, K. (1974). Failures in Upward Communication in Organizations: Three Possible Culprits. *The Academy of Management Journal*, *17(2),* 205-215

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. doi:10.1016/0022-1031(77)90049-x

Sharma, A.K. (2005), *Textbook of correlations and regression*. New Delhi: Discovery publishing house

Smith, A., & Fortunato, V. (2008). Factors Influencing Employee Intentions to Provide Honest Upward Feedback Ratings. *Journal Of Business And Psychology*, *22*(3), 191-207. doi: 10.1007/s10869-008-9070-4

Stoneman, K., Bancroft, E., & Hailing, C. (1995). Upward feedback for organizational change. *Performance + Instruction*, *34*(7), 12-17. doi: 10.1002/pfi.4170340705

Tourish, D., & Robson, P. (2004). Critical upward feedback in organisations: Processes, problems and implications for communication management. *Journal Of Communication Management*, *8*(2), 150-167. doi: 10.1108/13632540410807628

Tsui, A., & Barry, B. (1986). Research Notes: Interperso nal Affect and Rating Errors. *Academy Of Management Journal*, *29*(3), 586-599. doi: 10.5465/256225

van Dierendonck, D., Haynes, C., Borrill, C., & Stride, C. (2007). Effects of upward feedback on leadership behaviour toward subordinates. *Journal Of Management Development*, *26*(3), 228-238. doi: 10.1108/02621710710732137

Weaver, R. & Prelec, D. (2013). Creating Truth-Telling Incentives with the Bayesian Truth Serum, *Journal of Marketing Research, 50(3),* 289-302

Wellington, J. & Szczerbinski, M. (2007). *Research Methods for the Social Sciences*. London: Continuum International Publishing Group

Appendices

Appendix 1: Questionnaire

Survey Flow

Block Randomisation: Block Control and Block Treatment

Survey Content

Hello there,

I am Hai Binh Dang, a master student from Erasmus School of Economics. Thank you for participating in my thesis research on managerial behaviours.

The following questionnaire asks how you perceive your direct managers/supervisors. It will take approximately 5 minutes to accomplish. You are free to cancel your participation any time you wish.

By completing this survey, you will have a chance to receive 10 euros.

All data gathered will be completely anonymous, confidential and will not be disclosed to any party.

In case of any questions or concerns, please contact me via email 534159hd@student.eur.nl.

Part 1: Demographic information

Q2. Are you currently employed?

- Yes
- No

Display these questions if Are you currently employed? = Yes

Q4.1 Number of years working in your current organisation

- Less than 1 year
- 1-3 years
- 3-5 years
- More than 5 years

Q5.1 Size of your organisation:

- Small (less than 20 full-time employees)
- Medium (20 to 100 full-time employees)
- Large (more than 100 full-time employees)

Display this question if Are you currently employed? = No

Q3. Have you ever been employed?

- Yes
- No

Skip To: End of Block if Have you ever been employed? = No

Display these questions if Have you ever been employed? = Yes

Q4.2. Number of years working in your last organisation

- Less than 1 year
- 1-3 years
- 3-5 years
- More than 5 years

Q5.2 Size of your last organisation:

- Small (less than 20 full-time employees)
- Medium (20 to 100 full-time employees)
- Large (more than 100 full-time employees)

Part 2: Upward feedback

Control (a)

Kindly indicate whether you agree or disagree with the following statements.

Treatment (b)

For each complete answer, you will earn a **Truth Score**. Truth Scoring, recently invented by an MIT professor and published in the academic journal Science, rewards you for answering truthfully. Even though only you know if you really provide honest answers, people who tell the truth score higher overall. You are most likely to maximise your overall Truth Score if you answer every question truthfully. By *truthfully*, I mean consider each question carefully, answer honestly and take care to avoid mistakes. The person with the highest Truth Score will receive 10 euros as a reward.

Q6 Being dishonest will affect my Truth Score negatively.

- Yes
- No

Display this question if Being dishonest will affect my Truth Score negatively. = No

Q6.1 Answering the questions dishonestly will negatively affect your Truth Score.

Q8. The researcher can verify whether I lie or not.

- Yes
- No

Display this question if The researcher can verify whether I lie or not. = Yes

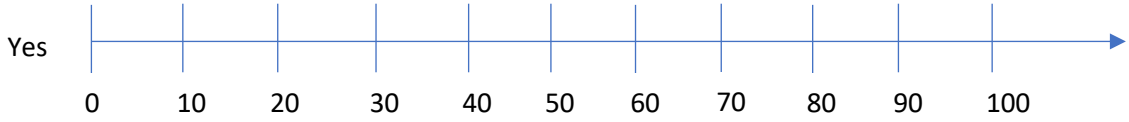
Q8.1 The researcher and everyone else cannot verify your truthfulness.

Q9. My manager provides constructive and actionable feedback that enables me to enhance my personal performance

A. Yes

B. No

Q9.1 Out of 100 respondents, how many do you think will answer “Yes” to the above question?

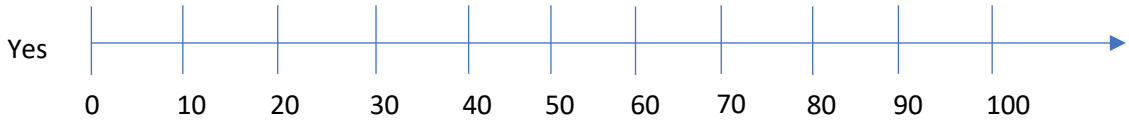


Q10. My manager gives me sufficient room of autonomy for decision making

A. Yes

B. No

Q10.1 Out of 100 respondents, how many do you think will answer “Yes” to the above question?

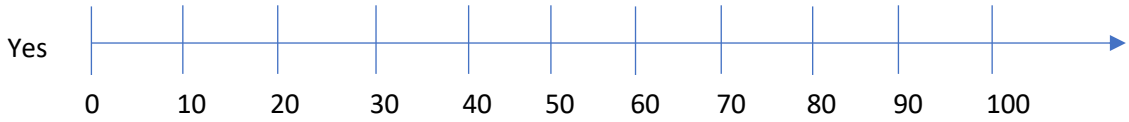


Q11. My manager shows necessary care for me as a person

A. Yes

B. No

Q11.1 Out of 100 respondents, how many do you think will answer “Yes” to the above question?

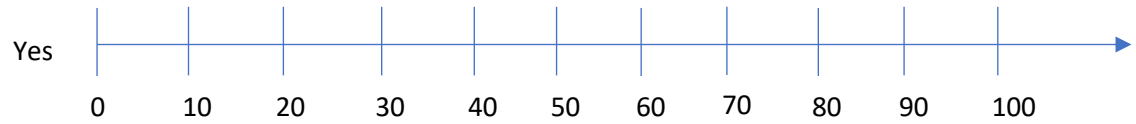


Q12. My manager demonstrates that she/he holds expertise in managing people effectively

A. Yes

B. No

Q12.1. Out of 100 respondents, how many do you think will answer “Yes” to the above question?

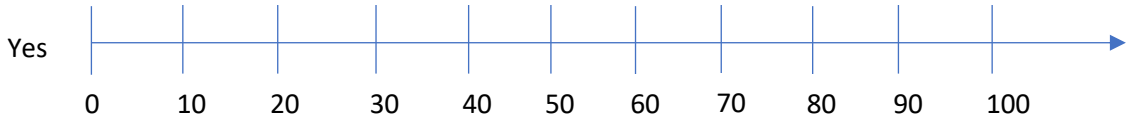


Q13. My manager clearly communicates the objectives and assign clear goals for each team member

A. Yes

B. No

Q13.1 Out of 100 respondents, how many do you think will answer “Yes” to the above question?

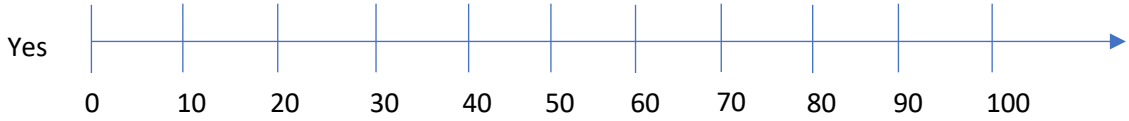


Q14. My manager constantly holds me and other subordinates accountable

A. Yes

B. No

Q14.1 Out of 100 respondents, how many do you think will answer “Yes” to the above question?

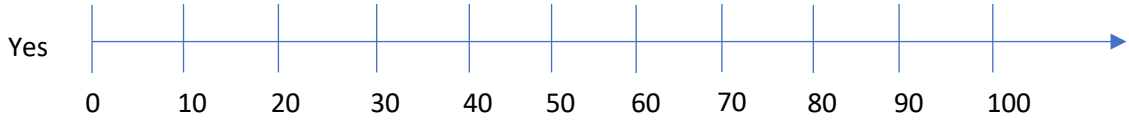


Q15. My manager effectively keeps the team focused on task completion even under pressure

A. Yes

B. No

Q15.1 Out of 100 respondents, how many do you think will answer “Yes” to the above question?



Part 3: Biases

Kindly indicate the extent to which you agree or disagree with the following statements

	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree
Q16. I will only give positive feedback toward manager with whom I have a good relationship					
Q17. I will be dishonest in giving feedback toward my manager as I fear that this can potentially harm my career					
Q18. I will be dishonest in giving feedback toward my manager as I feel that my feedback will not make any change to the practices of HR management of my company					
Q19. I will be dishonest in giving feedback toward my manager as I fear that the					

negative comment can potentially harm her/his career

Part 4: Incentive

Control:

By completing this survey, you will have a chance to receive 10 euros. Please leave your email below should you wish to take part in the random lottery.

Note: Providing your email address does not harm your anonymity as your answers are separated from it. Once the winner is selected, the email information will be deleted from the database.

Treatment:

Please leave your email below (you might be the winner!)

Note: Providing your email address does not harm your anonymity as your answers are separated from it. Once the winner is selected, the email information will be deleted from the database.

This is the end of the survey. Thank you for your kind contribution!

Appendix 2: Ethical Consideration

The first point regards with consent of participation, with participants taking part in the study on their complete willingness. To assure this willingness, the author employed a number of standard procedures including seeking for participant's consent through survey consent form or provision of participant's information sheet which provides basic Q&A about the research information. Even before the survey took place, questionnaire also recapped the information of the data collection process, such as the duration required for answering and other rights or risks so that the participants can make their owned decisions of participation. Respondents was also clearly informed that their participation was up to their choice and that they could cancel the survey session and thereby would quit from the research at any moment of their free will. With all these measures being employed, it was possible to assure that participation of all respondents was on their genuine of consent.

To assure the commitment to privacy protection, the researcher engages to making sure that any private information of the participants remains confidential and would not be disclosed to any third party without the participant's consent. This commitment is clearly presented in the introduction of the questionnaire while participants were also informed to not provide information of personal identity (i.e. name or address). Following that, the body of the research paper also respected this right by presenting the outputs of the data under lens of quantitative inference only, while none of information can be used to reflect personal identity.

For a human-involvement study without physical contact as this one, the only possible harm subjecting to the participants was mental harm for being exposed to the questions in questionnaire. To prevent this matter, the researcher conducted the questionnaire design with care and skills, avoiding any sensitive or provocative questions which may potentially arouse the mental distress. Likewise, the language of either questionnaire and research paper was also carefully edited to avoid offensive or discriminatory sense. Additionally, the selection of participants to the study was on a mixed basis of purposive sampling and snowball sampling, yet there was no difference in the way information is treated, considering the ignorance of respondent's personal identity.

With the consideration of the ethical issues in line with standard business ethics, it is possible to assure that the research could avoid the ethical violation; hence, research outputs are reliable and legitimate.

Appendix 3: Power Calculation

For testing optimality of the used sample sized, power calculation was employed through adoption of correlation test of the two samples. As a rule of thumb, an alpha significance level of 0.05 was assumed for the test, indicating that the error margin can be at 5%. Based on the two samples with sample size of $n1$ = 103 and $n2$ = 108, the correlation coefficients between upward feedback and bias was found at 0.53 and 0.14, respectively.

Estimated power for a two-sample correlations test

Fisher's z test

Ho: r2 = r1 versus Ha: r2 != r1

Study parameters:

alpha = 0.0500

N = 211

N1 = 108

N2 = 103

N2/N1 = 0.9537

delta = -0.3900

r1 = 0.5300

r2 = 0.1400

Estimated power:

power = 0.8953

To achieve this power, the required sample size will be:

. power twocorrelations 0.53 0.14, alpha(0.05) beta(0.2)

Performing iteration ...

Estimated sample sizes for a two-sample correlations test

Fisher's z test

Ho: r2 = r1 versus Ha: r2 != r1

Study parameters:

alpha = 0.0500

beta = 0.2000

delta = -0.3900

r1 = 0.5300

r2 = 0.1400

Estimated sample sizes:

N = 162

N per group = 81

As can be seen, it would require a sample of 81 in each group for achieving the desired power result. Therefore, it can briefly conclude that the sample of 103 and 108 observations in this study was sufficient to make meaningful and statistically reliable conclusion.

Appendix 4: Regression Models

Regress if BTS==0

Number of obs = 108

R-squared = 0.6757

F(4, 103) = 53.64

Source	SS	df	MS
Model	5.46386625	4	1.36596656
Residual	2.62286826	103	.02546474
Total	8.08673451	107	.075576958

Regress if BTS==1

Number of obs = 108

R-squared = 0.1801

F(4, 98) = 5.38

Prob > F = 0.0006

Source	SS	df	MS
Model	0.898227072	4	0.224556768
Residual	4.09008259	98	0.041735537
Total	4.98830967	102	0.048904997