

UDC 005.32:331.101.3:519.8

DOI [https://doi.org/10.15589/znп2026.1\(503\).2.24](https://doi.org/10.15589/znп2026.1(503).2.24)

BAYESIAN TRUTH SERUM FOR TRUE PERFORMANCE DIAGNOSTICS

БАЄСІВСЬКА СИРОВАТКА ПРАВДИ
ДЛЯ ПРАВДИВОЇ ДІАГНОСТИКИ ПРОДУКТИВНОСТІ

Andrii R. Yakymiv
yakymiv1995@gmail.com
ORCID: 0009-0000-1115-2905

А. Р. Якимів,
аспірант

Kyiv National University of Construction and Architecture, Kyiv
Київський національний університет будівництва і архітектури, м. Київ

Abstract. Purpose. The purpose of this study is to investigate the potential of the Bayesian Truth Serum mechanism for identifying hidden causes of changes in employee productivity using corporate survey data.

Methodology. The study applies the Bayesian Truth Serum mechanism to extract more reliable information from corporate survey responses. The method collects two signals from each respondent: a personal answer to a survey question and an estimate of how other participants are expected to respond. This structure allows analysis not only of individual opinions but also of respondents' perceptions of the collective distribution of answers. The analytical procedure compares the empirical frequency of each response option with the predicted frequency derived from respondents' expectations. Based on this comparison, an information score is calculated, reflecting the difference between observed and expected response distributions. Options that occur more frequently than predicted may indicate underestimated factors influencing productivity. To demonstrate the approach, a model corporate survey aimed at identifying causes of productivity changes was analyzed. Aggregated responses were used to calculate empirical frequencies, predicted frequencies, and information scores, allowing identification of patterns that diverge from expected collective perceptions.

Results. The results demonstrate that the Bayesian Truth Serum mechanism allows identification of responses that significantly differ from expected collective evaluations and therefore contain additional information about real factors influencing productivity. The approach reduces the impact of socially desirable responses and other systematic distortions commonly present in traditional surveys.

Scientific novelty. The scientific novelty of the study lies in the application of the Bayesian Truth Serum mechanism for analyzing hidden drivers of productivity changes within organizational environments.

Practical value. The proposed approach can be used to improve the reliability of corporate surveys and support more accurate identification of factors influencing employee productivity, which may assist managers in making better-informed organizational decisions.

Key words: Bayesian Truth Serum; employee productivity; causal surveys; information indicator; corporate analytics; prioritization of interventions.

Анотація. Мета. Метою дослідження є аналіз можливостей застосування механізму Bayesian Truth Serum (Баєсівська сироватка правди) для виявлення прихованих причин змін продуктивності працівників на основі даних корпоративних опитувань.

Методика. У дослідженні застосовано механізм Bayesian Truth Serum як інструмент отримання більш достовірної інформації з результатів опитувань. Підхід враховує не лише власні відповіді респондентів, але й їхні очікування щодо відповідей інших учасників. Метод ґрунтується на порівнянні фактичних частот відповідей із прогнозованими та розрахунку інформаційного показника, який дозволяє визначити відповіді, що містять найбільш інформативні оцінки. Для демонстрації підходу розглянуто модельну ситуацію корпоративного опитування, спрямованого на виявлення факторів, що впливають на зміни продуктивності.

Результати. Отримані результати показують, що застосування механізму Bayesian Truth Serum дозволяє ідентифікувати відповіді, які суттєво відрізняються від очікуваних колективних оцінок і містять додаткову інформацію щодо реальних факторів, які впливають на продуктивність. Запропонований підхід зменшує вплив соціально бажаних відповідей та інших систематичних викривлень, що часто виникають у традиційних опитуваннях.

Наукова новизна. Наукова новизна дослідження полягає у застосуванні механізму Bayesian Truth Serum для аналізу прихованих чинників змін продуктивності працівників у межах організаційного середовища.

Практична значимість. Практична значимість отриманих результатів полягає у можливості використання запропонованого підходу для підвищення достовірності корпоративних опитувань та більш точного виявлення факторів, що впливають на продуктивність працівників, що може бути використано при прийнятті управлінських рішень.

Ключові слова: Bayesian Truth Serum; продуктивність працівників; причинні опитування; інформаційний показник; корпоративна аналітика; пріоритезація управлінських втручань.

PROBLEM STATEMENT

Research on organizational performance frequently relies on subjective data sources, including employee surveys, interviews, self-reports, and peer or manager assessments. In organizational practice, such instruments are widely used to diagnose internal processes, identify barriers to effective work, and analyze potential causes of productivity changes. Surveys are particularly attractive because they allow organizations to collect feedback directly from employees about factors that are difficult to observe using purely quantitative indicators.

However, behavioral research has accumulated substantial evidence showing that subjective survey data are often affected by systematic distortions. Responses may be influenced by socially desirable behavior, impression management, or contextual pressures related to organizational hierarchy and expectations. Employees may therefore avoid expressing critical opinions about management practices or internal processes and instead provide answers that correspond to what they believe to be the expected or acceptable organizational narrative. As a result, surveys may reflect perceived norms rather than actual experiences of employees.

Another important limitation arises from common method bias, which occurs when both potential causes and outcomes are measured using the same questionnaire instrument. In such cases, relationships between variables may appear stronger than they actually are because the measures share the same response patterns. Consequently, corporate surveys may underestimate sensitive or inconvenient causes of productivity decline while overestimating explanations that are more socially acceptable or widely discussed.

In addition, employees' perceptions of collective opinion often differ from the actual distribution of responses within a group. When respondents attempt to predict how others will answer survey questions, these expectations may diverge from the real frequencies of responses. This gap between perceived consensus and actual opinions may contain important information about hidden organizational issues that remain undetected by traditional survey analysis.

These limitations indicate the need for methodological approaches that can extract more reliable signals from subjective survey data without relying on an external «correct answer». One such approach is the Bayesian Truth Serum mechanism, which evaluates responses not only by their observed frequency but also by comparing them with respondents' predictions about the answers of

others. This mechanism makes it possible to detect unexpectedly common responses and reveal hidden factors influencing employee productivity.

ANALYSIS OF LATEST RESEARCH

Research on organizational performance relies heavily on subjective data sources – self-reports, peer and manager ratings, surveys, interviews, and pulse surveys. At the same time, behavioral science has accumulated a significant body of evidence about systematic distortions of such data: from socially desirable responses and impression management to common method bias, where both causes (predictors) and outcomes are measured by the same questionnaire, making the relationship between them appear stronger than it actually is (correlations are artificially inflated). [5], [6]. For organizational practice, this means that even well-designed questionnaires and interviews risk underestimating «inconvenient» causes of performance decline or overestimating popular but superficial explanations; a significant body of work explicitly states the sensitivity of self-reports to context, sanctions, and management expectations [7], [8], [9]. It is against this background that mechanisms for encouraging truthfulness without an external «truth key» emerge – primarily Bayesian Truth Serum (BTS) and related peer-prediction approaches.

Prelec's original work showed that it is possible to award points not for the popularity of an answer, but for its «unexpected prevalence» relative to the respondents' own predictions of the distribution of others' answers. The information-score combination (log ratio of actual to predicted frequencies) deducts points for inaccurate predictions, so it is most profitable to tell the truth – even when the «correct» answer cannot be verified. [1]. Later, versions appeared that work on small samples and under milder conditions (even when participants do not have a «common basic opinion»): for example, robust BTS for «yes/no» questions [2]. For large-scale anonymous data collection (crowdsourcing, job evaluation, sensors), Peer Truth Serum (PTS) has been proposed, which simultaneously encourages both truthfulness and real effort on the part of the respondent [3]. In parallel, a «surprisingly popular» rule has emerged: choose the answer that occurs more often than people expected [4]. Together, this provides a set of tools that allows you to obtain truthful judgments without an external «correct» answer – exactly what is needed in corporate surveys, where the truth is often absent or comes late. When we try to find out the reasons for changes in productivity, data from company surveys is often distorted. If both «causes» and «results»

are collected in the same questionnaire, if there is no anonymity, and the questions are sensitive (about overload, conflicts, quality of management), people are more likely to underestimate negative signals and answer «as usual» [5], [6], [7]. Research distinguishes between two types of socially desirable responses: impression management and self-deception; in corporate surveys, the former is the most damaging, as responses can have direct or implicit consequences for the respondent [8]. And outside HR, the same is evident: meta-reviews show that socially desirable responses significantly degrade the quality of self-reports, so the use of mechanisms such as BTS helps to obtain more reliable indicators of causes at a large scale [9]. In parallel with BTS, a broader theoretical program of information-theoretic and peer-prediction mechanisms for non-verification detection has emerged, unifying BTS with previous approaches (Peer Prediction, Dasgupta-Ghosh) through payoffs proportional to the mutual information of participants' signals. This framework explains when and why «honesty» becomes the dominant strategy, particularly in multiple-question settings, and how to design payoffs to prevent collusion and «coordinated» responses [10]. For the practice of large companies, this opens the way to large-scale causal surveys: we can ask not only *what happened* (performance assessment), but also *why* (up/down factors), while stimulating truthfulness and identifying «unexpectedly common» reasons that conventional questionnaires often fail to detect.

SEPARATION OF PREVIOUSLY UNSOLVED PARTS OF THE PROBLEM

Despite the growing body of research on survey reliability and information elicitation mechanisms, several important issues remain insufficiently explored. Existing studies primarily focus on the theoretical properties of mechanisms such as Bayesian Truth Serum and its variations, as well as their applications in experimental settings, crowdsourcing environments, and prediction markets. However, the use of these mechanisms for analyzing corporate survey data aimed at identifying the causes of changes in employee productivity remains limited. In particular, the potential of Bayesian Truth Serum to detect hidden organizational factors that may be underestimated or misrepresented in traditional employee surveys has not been sufficiently investigated. As a result, there is still a lack of methodological approaches that allow organizations to reveal latent drivers of productivity changes using subjective survey data while reducing the influence of common response biases.

PURPOSE OF THE RESEARCH

The purpose of this study is to investigate the potential of the Bayesian Truth Serum mechanism for identifying hidden causes of changes in employee productivity based on corporate survey data.

To achieve this goal, the following tasks are set:

- to analyze existing approaches to collecting and interpreting subjective survey data in organizational research and corporate analytics;

- to examine the principles and mechanisms of Bayesian Truth Serum as a tool for extracting informative signals from survey responses;

- to develop an approach for applying the Bayesian Truth Serum mechanism to corporate surveys aimed at identifying factors influencing employee productivity;

- to demonstrate the application of the proposed approach using a model scenario of a corporate causal survey and analyze the obtained results.

METHODS, OBJECT AND SUBJECT OF RESEARCH

The object of the research is the process of identifying factors influencing employee productivity in organizational environments based on subjective survey data.

The subject of the research is the application of the Bayesian Truth Serum mechanism for analyzing corporate survey responses in order to reveal hidden factors affecting employee productivity.

The methodological basis of the study includes approaches from behavioral economics, information elicitation theory, and organizational analytics. In particular, the research relies on the Bayesian Truth Serum mechanism, which allows the extraction of informative signals from survey responses in situations where an objective «ground truth» is not directly observable.

The proposed approach is based on collecting two types of responses from each participant. First, respondents provide their own answer to a question concerning potential factors influencing productivity. Second, they estimate the expected distribution of responses among other participants. This structure allows the analysis to incorporate both individual opinions and respondents' expectations about collective judgments.

The analytical procedure involves calculating empirical response frequencies and comparing them with predicted frequencies derived from respondents' expectations. Based on this comparison, an information score is calculated for each response option. The score reflects the difference between observed and expected frequencies and makes it possible to identify responses that occur more often than predicted. Such responses may indicate underestimated or hidden factors influencing productivity.

To illustrate the proposed methodology, a model corporate survey scenario was analyzed. The collected responses were aggregated to compute empirical frequencies, predicted frequencies, and information scores for each response option. The results of this procedure provide a basis for identifying response patterns that differ from expected collective perceptions and may therefore indicate latent organizational issues affecting employee productivity.

MAIN PART

Model definition. In our case, each employee answers two blocks of questions. First, he chooses an answer to the question itself (for example, «what factor most influenced the productivity of this sprint?» or «how effective was the team on a scale of 1-5»). Then he makes a prediction:

in what proportions, in his opinion, other respondents will choose each of the options. That is, together with his own answer, the person provides a distribution of percentages for all options, which in total give 100%. It is this simple technique – «answer + prediction of others' answers» – that is, the basis of Bayesian Truth Serum (BTS) in Preletz's formulation [1]. Next, we summarize the data for the entire sample. For each answer option k , we calculate the real frequency x_k – that is, what proportion of people actually chose this option. Similarly, for each option, we calculate the «average prediction» y_k . Here, we take not the usual mean, but the geometric mean: it is convenient to think of it this way – we average the logarithms of the forecasts, and then «return» from the log scale back. The intuition is simple: the geometric mean is less «distorted» by extreme estimates and more honestly reflects the typical expectations of the group.

The key to BTS is to compare how often a variant actually occurs with how often respondents expected to see it. To do this, an information score is introduced:

$$Info = \log \frac{x_k}{y_k} \quad (1)$$

If $Info(k) > 0$, the option turned out to be «unexpectedly common» (more common than predicted). If $Info(k) < 0$, on the contrary, it was overestimated. A particular respondent receives exactly the information score corresponding to the option he chose. Thus, it is not the «popularity» but the «unexpected prevalence» of the answer that is rewarded.

We evaluate the accuracy of each person's prediction separately. The idea is simple: the closer his predicted distribution $y_{k,r}$ is to the actual x_k , the better the score; the further away, the greater the penalty. It is convenient to write this as

$$Pred(r) = a \sum_k x_k \log \left(\frac{y_{k,r}}{x_k} \right) \quad (2)$$

where a is the penalty weight (from 0 to 1). If the person guessed the distribution correctly, the sum is zero – this is the best possible result for this component. If the person consistently missed, the value will be negative. In practice, a specifies «how much» we value the accuracy of the prediction relative to the information score: $a = 1$ makes the game balanced within the group, $a < 1$ – mitigates the penalty for prediction errors.

The respondent's final score is the sum of two parts:

$$S_r = Info(ChosenAnswer) + Pred(r) \quad (3)$$

where $Info(ChosenAnswer)$ is the unexpected prevalence, $Pred(r)$ is how well you guessed the group.

This simple expression creates the necessary incentive: to have a high score, it is beneficial to answer honestly (because then you are more likely to be in the «unexpectedly truthful» minority) and to make realistic predictions about others (because points are deducted for mistakes). Even when the «correct» answer cannot be verified, the best strategy is to tell the truth and soberly evaluate the opinions of colleagues.

To calculate the formulas above, we will need intermediate formulas for individual values. Let us denote the answer options by the index $k \in \{1, \dots, m\}$ and the respondents by $r \in \{1, \dots, n\}$:

– The actual frequency of variant k :

$$x_k = \frac{1}{n} \sum_r 1\{respondentRpickK\} \quad (4)$$

This is a simple percentage: what proportion of people actually chose each reason.

– The «group forecast» for option k is the geometric mean of the individual forecasts

$$\log y_k = \frac{1}{n} \sum_{r=1}^n \log y_{k,r} \quad (5)$$

Next, this value must be normalized to 1. If someone's forecast is 0%, then a small constant (for example, 10^{-3}) is added before the logarithm and then normalized again. The geometric mean is important here: it «smooths out» single over/undershoots and better reflects the typical expectation of the group – that's how BTS is designed

Then, from these two aggregates, scores are calculated using formula (3). The first part rewards an «unexpectedly truthful» answer, the second part punishes unrealistic expectations about the group. When the respondent's prediction coincides with x , the second sum is zero – it can't get any better.

There are a few practical nuances. If someone had given zero percent for an option, the formulas would have $\log 0$ to avoid this, a tiny constant (e.g., 10^{-6}) would be added to the distributions and then normalized to 100%. For questions of the «1-5 scale» type, everything works the same, but the «options» are played by scores 1, 2, 3, 4, 5. For questions with a list of factors («what influenced performance?»), it is most convenient to ask to choose the main factor and separately give a forecast for all factors; then the formulas remain unchanged. If you want to allow several factors at once, you can reduce them to a sequence of simple choices (several sub-questions) or ask to distribute «100 influence points» between the options of the sample: these are the reasons that turned out to be «unexpectedly common.» In the context of performance, such signals are especially valuable: they often highlight a «wildcard» factor that management or a wider range of employees have underestimated. Combining these findings with our quantitative predictions (from previous sections), we get a closed loop: we predict when performance will change, and at the same time transparently reveal why it happened – at the scale of a large organization and without the need for expensive series of 1-on-1 interviews.

Experimental results. *Questions and answer format.* In a practical example, we model one sensitive question: «What is the main reason for the performance dip over the last 2 weeks (sprint)?» The respondent takes two steps: chooses one reason (his answer) and gives a forecast – in what proportions, in his opinion, others will distribute the answers between all options (the sum is 100%). This way we get both real choices and expectations about colleagues' choices – this is the key to BTS.

We work with 20 respondents and record a set of reasons: A – «unplanned tasks», B – «excessive meetings», C – «vague requirements», D – «tech debt», E – «slow CI/CD», F – «dependencies on other teams», G – «frequent interruptions/chat», H – «incidents in prod», I – «other». We make the predictions sparse: usually the respondent distributes the percentages between only 2-3 reasons (the others are given zeros). This is closer to real behavior – people do not keep in mind the complete distribution of all options.

BTS response format:

- choose one main reason;
- give a prediction of how you think others' answers will be distributed (shares for all options, sum = 100%).

What kind of scenario are we showing? To demonstrate the analytical utility of BTS, we consider two data states:

1. Basic sparse state: we generated individual responses and forecasts with typical biases (for example, people intuitively overestimate the role of «meetings» or «interruptions»), calculated and obtained a «natural» through the geometric mean.

2. Scenario of «inflated expectations for meetings»: we left the actual frequencies unchanged (i.e., people's real elections did not affect them), but formed a target group forecast.

To make the geometric mean truly equal to these values, all respondents were given the same (or nearly the same) forecast line. The goal is to show a case where the corporate narrative expects meetings to dominate (80%) while the actual election x_k does not confirm this. In this case, $Info(meetings)$ becomes negative (overestimation), and the underestimated causes with non-zero and low receive high positive Info and come to the fore as hidden drivers of subsidence.

This is an important point of the methodology: BTS does not require an external «right answer». It reveals the gap between what people actually chose x_k and what they expected from each other y_k . It is this gap that is a marker of «unexpected prevalence» or «overestimation» of reasons. In large companies, this helps filter out loud but exaggerated topics and highlight those that really «hurt», although they sound less often.

Results. Information for respondents.

Table 1. Respondents' answers

[R]	[O]	A	B	C	D	E	F	G	H	I
R01	A	10	25	0	31	0	0	0	34	0
R02	A	0	54	0	0	0	16	0	0	30
R03	A	41	0	21	0	38	0	0	0	0
R04	B	59	28	0	0	0	0	13	0	0
R05	B	0	92	0	0	0	0	0	0	8
R06	B	10	64	0	0	0	0	0	0	26
R07	C	56	44	0	0	0	0	0	0	0
R08	C	0	63	0	0	0	37	0	0	0
R09	D	0	0	66	4	0	0	30	0	0
R10	D	23	0	0	0	18	33	0	26	0
R11	D	0	65	0	0	0	35	0	0	0
R12	E	0	0	11	14	0	42	0	0	33
R13	F	49	0	0	51	0	0	0	0	0
R14	F	29	45	0	26	0	0	0	0	0
R15	F	0	33	33	16	0	18	0	0	0
R16	F	0	59	0	0	0	22	19	0	0
R17	G	0	0	0	26	0	0	0	8	66
R18	G	0	0	0	8	0	0	0	52	40
R19	H	32	0	0	40	17	11	0	0	0
R20	I	14	86	0	0	0	0	0	0	0

Table 2. Scoring of respondents

[R]	[O]	Info	Pred	Score
R01	A	1.1774	-12.5125	-11.3352
R02	A	1.1774	-15.0116	-13.8342
R03	A	1.1774	-17.5944	-16.4171
R04	B	-1.835	-14.9653	-16.8003
R05	B	-1.835	-20.1575	-21.9925
R06	B	-1.835	-16.3532	-18.1882
R07	C	8.673	-17.4665	-8.7935
R08	C	8.673	-16.1432	-7.4702
R09	D	2.659	-16.5332	-13.8743
R10	D	2.659	-13.7035	-11.0445
R11	D	2.659	-16.1507	-13.4917
R12	E	9.3205	-12.4761	-3.1556
R13	F	4.1363	-17.4643	-13.328
R14	F	4.1363	-13.6185	-9.4822
R15	F	4.1363	-9.8611	-5.7248
R16	F	4.1363	-13.6618	-9.5256
R17	G	10.0319	-18.9882	-8.9563
R18	G	10.0319	-19.0969	-9.065
R19	H	7.9955	-11.183	-3.1875
R20	I	5.3233	-17.5759	-12.2526

Table 3. Summary by options

Reason	x_k	y^k %	Info
Frequent interruptions/chatting	10.0	0.0 %	10.0319
Slow CI/CD	5.0	0.0 %	9.3205
Unclear requirements	10.0	0.0 %	8.673
Incidents in production	5.0	0.0 %	7.9955
Other	5.0	0.0 %	5.3233
Dependencies on other commands	20.0	0.3 %	4.1363
Tech debt	15.0	1.1 %	2.659
Unplanned tasks	15.0	4.6 %	1.1774
Excessive meetings	15.0	94.0 %	-1.835

Due to the strong collective bias of expectation towards «excessive meetings», *Pred* is on average noticeably negative (mean ≈ -15.53 ; median -16.15 ; range -20.16 to -9.86). This is normal: when forecasts do not match the actual x , there is a penalty for inaccuracy.

– Distribution. Minimum = -21.9925 , median = -11.1898 , 75th percentile = 8.9156 , maximum = 3.1556 , mean ≈ -11.3960 . So, some participants get high x scores due to «unexpectedly common» answers (large positive *Info*), others get low scores due to «overestimated» (negative *Info*) or «moderate» reasons.

– Respondents with good answers.

○ R12 («Slow CI/CD»): *Info* ≈ 9.3205 , *Pred* ≈ -12.4761 , *S* ≈ -3.1556 . – the best result.

○ R19 («Incidents in Prod»): *Info* ≈ 7.9955 , *Pred* ≈ -11.1830 , *S* ≈ -3.1875 .

○ R15 («Dependencies on other commands»): *Info* ≈ 4.1363 , *Pred* ≈ -9.8611 , *S* ≈ -5.7248 .

– Respondents with worse answers.

○ R05 («Exc. meetings»): *Info* ≈ -1.8350 , *Pred* ≈ -20.1575 , *S* ≈ -21.9925 .

○ R06 («Exc. meetings»): *Info* ≈ -1.8350 , *Pred* ≈ -16.3532 , *S* ≈ -18.1882 .

○ R04 («Exc. meetings»): *Info* ≈ -1.8350 , *Pred* ≈ -14.9653 , *S* ≈ -16.8003 .

These respondents have both a «negative» *Info* (because the meetings are overestimated) and inaccurate forecasts relative to the actual x , so the sum gives a very low *S*.

What answers were chosen more often? «Dependencies on other teams» – 4/20 (20%), «Unplanned tasks» – 3 (15%), «Excessive meetings» – 3 (15%), «Tech debt» – 3 (15%), «Vague requirements» – 2 (10%), «Frequent interruptions / chat» – 2 (10%), «Slow CI/CD» – 1 (5%), «Incidents in prod» – 1 (5%), «Other» – 1 (5%). This gives an intuition for the actual choices, but the decisive factor for *S* remains the *Info* of the selected option plus the *Pred* of a specific person.

DISCUSSION OF RESULTS

We estimate three values for each reason: the actual share of elections (x_k), group expectations (y_k) and the indicator.

$$Info = \log \frac{x_k}{y_k}$$

The interpretation is straightforward: if *Info* > 0 , the reason turned out to be unexpectedly wide-spread (the group underestimated it); if *Info* < 0 , it is

overestimated at the level of the entire sample. In our scenario, «Excessive meetings» has ($x = 15\%$) versus ($y = 80\%$), so (*Info* $= -1.674$). That is, the narrative «meetings are the main culprit» turned out to be overestimated. On the other hand, «Dependencies on other teams» ($x = 20\%$ versus $y = 5\%$, *Info* $= 1.3863$), «Frequent interruptions / chat» ($x = 10\%$ versus $y = 5\%$, *Info* $= 0.6931$) and «Unplanned tasks» ($x = 15\%$ versus $y = 10\%$, *Info* $= 0.4055$) are underestimated *silent culprits*: they occurred more often than expected. Additionally, we see a group of reasons that completely «fell off the radar» of expectations: for options with a given $y_k = 0$, any non-zero x_k gives a very large positive *Info* (we just technically add a small in front of the logarithm). Thus, «Tech debt», «Vague requirements», as well as «Slow CI/CD», «Incidents in prod» and «Other» indicate that these factors were not expected, but are actually present in the answers – this is a strong signal for further investigation.

From a management perspective, it is advisable to single out 2-3 causes with the highest positive *Info* as first-line candidates for action (process changes, tools, policies) in the final report, and to name 1-2 causes with the lowest *Info* as overestimated at the level of the entire sample. For both groups, it is worth making local slices (teams, roles, locations) and adding an impact assessment to prioritize not only by prevalence, but also by expected damage: this way we focus efforts on causes that both occur frequently and have the greatest impact on productivity.

It is important to note that a negative *Info* does not mean that the «cause is not important». It just means that in our sample it occurred less frequently than expected. For rare but potentially highly damaging causes, we always check subgroups and objective metrics (e.g., real-time in meetings, CI/CD queues, number of integration blockers, etc.).

CONCLUSIONS

The study demonstrates that the Bayesian Truth Serum mechanism can be effectively integrated into corporate causal surveys to diagnose changes in employee productivity in situations where no external «truth key» exists. By comparing respondents' actual choices with their expectations about others' answers, the method reveals discrepancies between perceived and real collective opinions. Such differences allow identification of underestimated causes of productivity decline as well as

topics that may be overestimated due to organizational narratives or socially desirable responses.

The demonstration scenario confirmed the presence of expectation bias in survey data: some widely discussed factors appeared less influential than expected, while several less visible causes turned out to be unexpectedly widespread. This property makes the approach useful for prioritizing managerial interventions by focusing attention on factors where observed responses systematically exceed expectations.

The limitations of the presented example include the relatively small sample size, the use of a single survey wave, and the assumption of one primary cause per respondent. Future research should focus on longitudinal applications of the approach, the development of more robust variants of the mechanism for small subgroups, the inclusion of multiple response options with influence weights, and the integration of Bayesian Truth Serum with predictive productivity models.

REFERENCES

- [1] Prelec, D. (2004). A Bayesian Truth Serum for Subjective Data. *Science*, 306(5695), 462–466. <https://doi.org/10.1126/science.1102081>
- [2] Witkowski, J., & Parkes, D. C. (2012). A Robust Bayesian Truth Serum for Small Populations. In *Proceedings of the AAAI Conference on Artificial Intelligence*. <https://ojs.aaai.org/index.php/AAAI/article/view/8261>
- [3] Radanovic, G., Faltings, B., & Jurca, R. (2016). Incentives for Effort in Crowdsourcing Using the Peer Truth Serum. *ACM Transactions on Intelligent Systems and Technology*, 7(4), Article 48. <https://doi.org/10.1145/2856102>
- [4] Prelec, D., Seung, H. S., & McCoy, J. (2017). A solution to the single-question crowd wisdom problem. *Nature*, 541, 532–535. <https://doi.org/10.1038/nature21054>
- [5] Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://pubmed.ncbi.nlm.nih.gov/14516251/>
- [6] Keiser, N. L., & Payne, S. C. (2019). Are employee surveys biased? Impression management as a response bias in workplace safety constructs. *Safety Science*, 118, 453–465. <https://doi.org/10.1016/j.ssci.2019.05.051>
- [7] Huebner, L. A., et al. (2021). Following Up on Employee Surveys: A Conceptual Framework for Organizing Action Planning. *Frontiers in Psychology*, 12, 733819. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.801073/full>
- [8] Zerbe, W. J., & Paulhus, D. L. (1987). Socially Desirable Responding in Organizational Behavior: A Reconceptualization. *Academy of Management Review*, 12(2), 250–264. <https://doi.org/10.5465/amr.1987.4307820>
- [9] Latkin, C. A., Edwards, C., Davey-Rothwell, M. A., & Tobin, K. E. (2017). The relationship between social desirability bias and self-reports of health behaviors. *Journal of Addiction and Prevention*, 8(1), 1–9. <https://pubmed.ncbi.nlm.nih.gov/articles/PMC5519338/>
- [10] Kong, Y., & Schoenebeck, G. (2016). An Information Theoretic Framework for Designing Information Elicitation Mechanisms That Reward Truth-telling. *arXiv preprint*. <https://arxiv.org/abs/1605.01021>

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] Prelec, D. (2004). A Bayesian Truth Serum for Subjective Data. *Science*, 306(5695), 462–466. <https://doi.org/10.1126/science.1102081>
- [2] Witkowski, J., & Parkes, D. C. (2012). A Robust Bayesian Truth Serum for Small Populations. In *Proceedings of the AAAI Conference on Artificial Intelligence*. <https://ojs.aaai.org/index.php/AAAI/article/view/8261>
- [3] Radanovic, G., Faltings, B., & Jurca, R. (2016). Incentives for Effort in Crowdsourcing Using the Peer Truth Serum. *ACM Transactions on Intelligent Systems and Technology*, 7(4), Article 48. <https://doi.org/10.1145/2856102>
- [4] Prelec, D., Seung, H. S., & McCoy, J. (2017). A solution to the single-question crowd wisdom problem. *Nature*, 541, 532–535. <https://doi.org/10.1038/nature21054>
- [5] Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://pubmed.ncbi.nlm.nih.gov/14516251/>
- [6] Keiser, N. L., & Payne, S. C. (2019). Are employee surveys biased? Impression management as a response bias in workplace safety constructs. *Safety Science*, 118, 453–465. <https://doi.org/10.1016/j.ssci.2019.05.051>
- [7] Huebner, L. A., et al. (2021). Following Up on Employee Surveys: A Conceptual Framework for Organizing Action Planning. *Frontiers in Psychology*, 12, 733819. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.801073/full>
- [8] Zerbe, W. J., & Paulhus, D. L. (1987). Socially Desirable Responding in Organizational Behavior: A Reconceptualization. *Academy of Management Review*, 12(2), 250–264. <https://doi.org/10.5465/amr.1987.4307820>
- [9] Latkin, C. A., Edwards, C., Davey-Rothwell, M. A., & Tobin, K. E. (2017). The relationship between social desirability bias and self-reports of health behaviors. *Journal of Addiction and Prevention*, 8(1), 1–9. <https://pubmed.ncbi.nlm.nih.gov/articles/PMC5519338/>
- [10] Kong, Y., & Schoenebeck, G. (2016). An Information Theoretic Framework for Designing Information Elicitation Mechanisms That Reward Truth-telling. *arXiv preprint*. <https://arxiv.org/abs/1605.01021>

© Якимів А. Р.

Дата першого надходження статті до видання: 12.02.2026

Дата прийняття статті до друку після рецензування: 10.03.2026

Дата публікації (оприлюднення) статті: 22.05.2026



Стаття поширюється на умовах ліцензії відкритого доступу (CC BY 4.0)