

# 回答品質の評価に対する相互作用アプローチ ーベイジアン自白剤の適用可能性についての検討ー

An Interactionist Approach to Evaluate Quality of Survey Response:  
Validating Bayesian Truth Serum in Consumer/Employer Surveys

小野 滋  
Shigeru Ono

1. はじめに
2. 回答品質への2つのアプローチ
3. ベイジアン自白剤とその実証研究
4. 実験
5. 考察

## 〈要旨〉

近年の調査手法研究では、ゲーム理論に代表される、調査を相互作用的行為として捉えるアプローチが大きな成果を挙げはじめている。その一つであるベイジアン自白剤(BTS)は、調査回答の真実性を評価・改善するための手法として注目される。しかし、BTS についての実証研究はいまだ十分とはいえない。本研究では、マーケティング・リサーチの文脈における BTS の適用可能性を検討した2つの実験について報告する。結果はBTS の実務における有用性を示唆した。調査手法研究に対する本研究の意義、ならびに今後の研究課題について論じる。

Recent literature on survey methodology has begun to adapt new survey viewpoints, especially in terms of game theory. The Bayesian Truth Serum (BTS) is one such new method derived from these interactionist perspectives. Scholars have argued that BTS provides a new way to evaluate and improve truthfulness of survey responses. However, despite the method's appeal, evidence to support these arguments is insufficient. Using two experimental studies, this paper demonstrates the performance of BTS in consumer/employee surveys in marketing research. The results support the applicability of BTS in marketing research practices, and its implication for research methodology and future directions are discussed.

## 1. はじめに

多種多様なデータが溢れる現代においても、質問紙調査は社会的意思決定を支援するための重要な手段の一つであり続けている。社会科学、ジャーナリズム、マーケティングなどさまざまな領域において、質問紙調査は人々の知識・態度を把握するための主要な手段の一つとなっている。よりよい調査結果を得ることは、多くの人々にとっての重大な関心事である。

調査法改善についての議論は、調査対象者（誰に訊くか）と調査方式（どのように訊くか）の二つの側面に大別できる。本論文は後者に焦点を当て、回答品質の評価と改善のための手法として近年注目を集めているベイジアン自白剤（Bayesian Truth Serum; BTS）についての実証実験を報告し、その有用性について議論する。

本研究の主たる関心は、マーケティング意思決定支援のための消費者調査・従業員調査の文脈における調査回答の品質改善にある。しかし、本研究が扱うベイジアン自白剤それ自体は、調査結果が活用される文脈に依存しない手法である。したがって、本研究における知見も、社会調査・選挙調査を含むさまざまな領域における調査の改善に対して含意を持つものと考えられる。

本報告の構成は次の通りである。まず回答品質の評価・改善に関するこれまでのアプローチを整理し（2章）、ベイジアン自白剤ならびに関連する実証研究について概観する（3章）。次に、マーケティング・リサーチにおけるベイジアン自白剤の適用可能性を検証するために行った2つの実験について報告する（4章）。最後に、2つの実験から得られた示唆と、今後の課題について考察する（5章）。

## 2. 回答品質への2つのアプローチ

### 2-1. 伝統的アプローチ

調査回答の品質をめぐる従来の議論を、回答の品質の評価に関わる問題と、回答の品質の改善に関わる問題に分けて整理してみよう。

従来、回答品質の評価という問題は、主に次の3つの観点から論じられてきた。

- (1) **回答の信頼性**. たとえば、複数回の調査に参加したある回答者が、同じ回答を示すはずの項目において実際に同じ回答を示したとき、それは信頼性の高い回答とみなされる。
- (2) **回答の妥当性**. たとえば、なんらかの外的基準に照らして当然に期待される回答と実際

の回答が一致したとき、それは妥当性の高い回答とみなされる。

- (3) **パラ・データの特性**. たとえば、回答にかかる時間が相対的に短かった回答や長かった回答には、疑いが向けられることが多い。

回答品質の向上という問題は、ほとんどの場合、調査主体が参加者に提示する情報の改善という観点から論じられてきた。すなわち、調査依頼の改善、教示の改善、調査項目の改善などである。そこでの一般的指針は、調査回答にバイアスを与えないこと、そして回答における認知的負荷を低く抑えること、であった。

これらの議論においては、調査主体が関心を持つなんらかの知識・態度を、調査回答者が（調査参加という行為に先行して）保持していることが前提となっている。回答の品質とは、調査回答者が保持しているそれらの知識・態度を、回答が正確に反映していることに他ならない。

### 2-2. 相互作用的アプローチ

一方、近年の調査法研究においては、こうした伝統的なアプローチとは異なる一連のアプローチが登場している。その代表例として、ゲーム理論の観点からの調査法研究が挙げられる。

ゲーム理論の観点から生まれた調査法研究の古典的な例として、Beckerらが提案したBDMメカニズム（Becker, DeGroot, & Marschak, 1964）がある。この手法では、設問への回答がその後の報酬と精密に関連づけられており、回答者からみると、正直かつ真剣な回答が報酬の最大化をもたらす仕組みとなっている（インセンティブ整合的メカニズム）。現在、BDMメカニズムは実験経済学における支払意思額の測定方法として広く用いられているほか、マーケティング・リサーチにおいても応用されている（Ding, 2007）。

ゲーム理論の観点に立つ調査法研究は、調査を調査主体と調査回答者との社会的相互行為として捉える。それはいわば情報とインセンティブの交換である。調査主体は、関心の対象となる事柄について調査回答者がなんらかの価値ある私秘的情報を開示しようと期待し、調査回答者にそれを開示させようとする。いっぽう調査回答者は受動的な測定対象ではなく、調査という社会的相互作用に関わる能動的な主体であり、情報の開示（すなわち回答）という行為を通じて、なんらかの社会的目的を達しようとする。調査主体にできること

		対象者A	対象者B	...
回答	Q1. この製品を買ってみたいですか？	はい	いいえ	...
回答率	Q2. Q1に「はい」と答える人は何パーセントいると思いますか？	25%	30%	...
→				
集計結果				
	実際の「はい」回答率	実際の「いいえ」回答率		
	25%	75%		
	「はい」回答率予測の幾何平均	「いいえ」回答率予測の幾何平均		
	31%	65%		
←				
	「はい」回答者に与える情報スコア	「いいえ」回答者に与える情報スコア		
	$\log(0.25/0.31)$	$\log(0.75/0.65)$		
	=-0.22	=+0.08		
	情報スコア	-0.22	+0.08	...
	「はい」についての対象者の回答率予測と、実際の回答率との比の対数	$\log(0.25/0.25)$	$\log(0.30/0.25)$	...
		=0.00	=+0.18	...
	「いいえ」についての対象者の回答率予測と、実際の回答率との比の対数	$\log(0.75/0.75)$	$\log(0.7/0.75)$	...
		=0.00	=-0.07	...
	上記の2つの値を実際の回答率で重みづけた和	$0.25 \times (0.00)$ $+ 0.75 \times (0.00)$	$0.25 \times (+0.18)$ $+ 0.75 \times (-0.07)$	...
		=0.00	=-0.01	...
	予測スコア	0.00	-0.01	...
	(情報スコア) + $\alpha$ (予測スコア)	-0.22+0.00	-0.08+(-0.01)	...
	(ここでは $\alpha=1$ の場合を示す)	=-0.22	=+0.07	...
	BTSスコア	-0.22	+0.07	...

図1. BTSスコアの算出例 (架空例)

は、調査回答という行為を通じて調査回答者が受け取る有形無形のインセンティブに注目し、調査回答者から見て正直かつ正確な回答が利益を最大化する回答となるよう、インセンティブの構造を設計することである。

このように調査を社会的相互作用として捉えなおすことで、調査品質についての議論も新たな地平を獲得する。この立場からみると、回答が表現している知識・態度とは、日常のコミュニケーションにおける意図がそうであるように、調査という行為に際して回答者がアドホックに生成したものである。回答の品質という問いは、「回答が調査参加者の知識・態度を正確に反映しているか」ではなく、むしろ、調査結果が社会現象の理解や意思決定支援に寄与するかどうかという問いへと形を変える。回答品質の向上のために調査主体が目指すべきは、バイアス源や認知的負荷を取り除くことというよりもむしろ、期待される情報が開示されるよう、調査という相互作用を最適に設計することであるということになる。

### 3. ベイジアン自白剤とその実証研究

#### 3-1. ベイジアン自白剤とは

近年では、ゲーム理論の観点から生まれた新し

い調査手法として、心理学者 Prelec (1994)が提案したベイジアン自白剤が注目を集めている。

ベイジアン自白剤は任意のカテゴリカル設問への回答に対するスコアリング・ルールである。調査参加者に設問への回答を求めるとともに、「その設問に他の人がどう答えるか」を予測する設問への回答を求める。設問例を下に示す。

問1. この製品を買ってみたいですか？ (はい / いいえ)
問2. 問1に「はい」と答える人は、何パーセントいると思いますか？ (___%)

対象者  $i$  ( $= 1, \dots, N$ ) のカテゴリ  $k$  ( $= 1, \dots, K$ ) に対する選択有無を  $x_{ik}$ 、他の人がカテゴリ  $k$  を選択する確率の予測についての回答を  $y_{ik}$  とし、それぞれの平均を

$$\bar{x}_k = \left(\frac{1}{N}\right) \sum_i x_{ik}, \quad \log \bar{y}_k = \left(\frac{1}{N}\right) \sum_i \log y_{ik}$$

とする。BTSスコアを下式で定義する。

$$BTS\ Score_i = \sum_k x_{ik} \log \frac{\bar{x}_k}{\bar{y}_k} + \alpha \sum_k \bar{x}_k \log \frac{y_{ik}}{\bar{x}_k}$$

上式右辺の第1項を**情報スコア**、第2項から $\alpha$ を取り除いた部分を**予測スコア**と呼ぶ。係数 $\alpha (> 0)$ はBTSスコアの性質を調整する定数であり、 $\alpha = 1$ のとき、全回答者を通じたBTSスコアの合計は0となる。

例として、二択型質問の架空のデータとBTSスコアの計算例を示す(図1)。BTSの核心となるのは情報スコアである。この例では、「いいえ」回答率の予測は平均して65%だが、実際の「いいえ」回答率は75%であった。このように、「みんなの予測より普通な」回答に高い情報スコアが与えられる。

Prelecらは、BTSスコアを回答の真実性(truthfulness)を表す指標として捉えることを提案している。その理由は直感的にはわかりにくいだが、Prelec(2004)は次のような形で説明している。人は他人の意見について推測する際、それが自分の意見に近いと推測する傾向がある。ある製品を買いたいと思っている人はみんなもそう思っているだろうと考える。その人たちからみて全員による「はい」回答率予測は低すぎる。同様に、ある製品を買いたくないと思っている人からみて全員による「いいえ」回答率予測は低すぎる。つまり、ある人の真の意見は「みんなが思うよりも普通な」回答になりやすい。

さらにPrelecは、ある自然な仮定の下でBTSスコアの期待値が真実申告において最大となり、BTSスコアの最大化のためには真実申告がベジアン・ナッシュ均衡解となることを示している。つまり、BTSスコアはインセンティブ整合的メカニズムを提供する。高いBTSスコアを得る方法は正直かつ真剣に答えることであり、他に方法はない。

### 3-2. ベイジアン自白剤に関する実証研究

ベイジアン自白剤はいくつかのゲーム理論的前提に基づくモデルに依拠しており、実際の挙動については実証的検討が必要である。1994年の提案以来、ベイジアン自白剤についての実証研究は、数は多くないものの着実に進められてきた。それらの研究は大きく二つの方向に分かれる。

#### 3-2-1. BTSによる回答品質の評価

一つは、BTSスコアを回答品質の評価手法として捉える方向である。

Howie, Wang, & Tsai(2011)は医師を対象とした調査で、回答のBTSスコアと行動との関係を調べている。医薬品の新製品を上市する前、製薬会社は販売予測のために医師を対象にした調査を行うことが多い。しかし、上市前に医師に聴取した処方意向と、上市後に医師がそれを実際に処方する回数は、通常大きくずれる。HowieらはBTSスコアで医師に重みをつけることで、実際の処方行動を処方意向で予測するモデルが説明力を増したと報告している。

Miller, Brailey, & Kirlik(1994)はデザイン教育評価において、学生に既存デザインについての設問への回答を求めたとき、(教員からみた回答の正しさではなく)その回答の情報スコアが、デザイン創造課題での成績を予測したと報告している。

#### 3-2-2. BTSによる回答品質の改善

もう一つの方向は、BTSスコアを調査参加報酬と連動させ、インセンティブ整合性の導入を試みるという方向である。参加者は報酬の最大化を目指して正直かつ真剣に回答するようになると期待される。

Weaver & Prelec(2013)は市場調査を想定した実験を行っている。実験参加者にブランド名などの名称を提示し、「知っている」と答えたら定額を支払うという試行では、参加者は実在しない名称に対しても「知っている」と答えやすくなる。そこで別の参加者には、質問にBTSを組み込み、「知っている」ときに定額を支払うだけでなく、情報スコアと比例した小銭も毎回支払った。試行を繰り返すと、参加者は実在しない名称を正しく「知らない」と答えるようになった。情報スコアと連動した報酬によって参加者の正直さが取り戻せたわけである。

John, Lowenstein, & Prelec(2012)は研究不正についての調査においてBTSを活用している。彼らは米国の心理学者を対象に、研究上の軽微な不正行為についての経験を聴取する調査を行った。調査参加への金銭的報酬のかわりに、参加報酬にあたる金額を調査者が慈善団体などに寄付すると約束した。回答は匿名で収集されるものの、回答者には自らの不正行為を正直に申告する動機づけが乏しい。そこで彼らは、一部の回答者に対して、調査参加に伴って生じる寄付の金額を、回答のBTSスコアと連動させると教示した。その結果、

この教示を行った群のほうが、不正行為を告白する割合が高くなった。

近年では、クラウド・ソーシングにおける報酬を BTS スコアと連動させる試みも行われている (Show, Horton, Chen, 2001; Frank, Cebrian, Pickard, Rahwan, 2017)。実験により効果にばらつきがあるものの、BTS の導入により、設問に対してワーカーが申告する回答の分布が正直かつ真剣な方向へと変動することが示されている。

#### 4. 実験

本報告では、マーケティング・リサーチにおける BTS の適用可能性を検証するために行った、ふたつのフィールド実験について報告する。

前節で概観したように、BTS についての実証研究は広がりを見せているが、いまだ十分とはいえない。とくに、消費者調査・従業員調査における回答品質の評価という文脈において、実務状況に近いフィールド実験は管見の限り見当たらない。

マーケティングの領域では、意思決定支援のための消費者調査や従業員調査が頻繁に行われており、調査自体が一つのビジネス領域をなしている。しかし、伝統的調査手法の減少とネット調査のコモディティ化に伴い、マーケティング実務家の間では、調査回答の品質への疑念、ひいては調査そのものへの疑念が高まっている。この領域での BTS の適用可能性を検証することは、調査実務家にとっても、またマーケティング実務家にとっても大きな意味を持っている。

##### 実験1：販売スタッフによる上市前販売予測

本実験では、従業員による上市前販売予測という文脈において、BTS の適用可能性を検討する。

**【背景】** メーカーは、新しい製品やプロモーションに対する顧客の反応を、実際の上市・実施前に知り、製品・プロモーションの改善や生産管理に活かしたいという強い動機づけを持っている。しかし、財の性質や秘匿性の観点から、上市・実施前の顧客調査が困難である場合も少なくない。こうした場合、顧客と日常的に接する販売スタッフ

からの意見集約が試みられることになる。

一方、従業員に対する調査には深刻な障害がある。たとえば、店頭販売員が本部スタッフから調査への参加を求められ、その調査のなかで、来期の新製品を提示され「どのくらい売れるでしょうか」と問われたとしよう。販売員の立場からみると、この設問に対して悲観的な回答を行うことは、モチベーションが低いと解される危険を伴う。このように、従業員から得た調査回答は強い評価懸念バイアスを受けている可能性が高い。

本実験では、このような文脈における BTS の適用可能性について検証する。BTS スコアが回答の真実性の指標となっているならば、強い評価懸念バイアスを受けていると思われる回答において、BTS スコアは低いはずである。

**【方法】** ある消費財メーカーによる、全国の販売員への定例調査のなかで実験を行った。この調査は勤務時間中に業務の一環としてなされている。調査は、自社の上市前新製品・上市済製品についてのいくつかの設問を含んでいる。

調査は匿名で行われると教示されており、実際に回答は匿名化した状態で収集されている。しかし、そのことは回答者から見て明らかではない。したがって、教示にも関わらず、回答者は回答の匿名性に疑念を持っている可能性がある。

なお、この製品カテゴリにおいては、販売員は特定のメーカーに専属する。新製品の上市前には、既存顧客に対する新製品の紹介活動が行われることが通例である。組織の構造上、個々の販売員の店頭での努力や上市前紹介活動の正確な実態について調査主体(本社)が把握するのは困難であり、販売員もそのことを理解している。回答した販売員の経験年数などは明らかでないが、一般に、販売員の製品知識は概して高く、また顧客についても豊かな知見を持っている。

全国の販売員 341 名から回答を収集した。本報告では回収率を明らかにできないが、調査参加は業務の一環であり、回収率は高い。

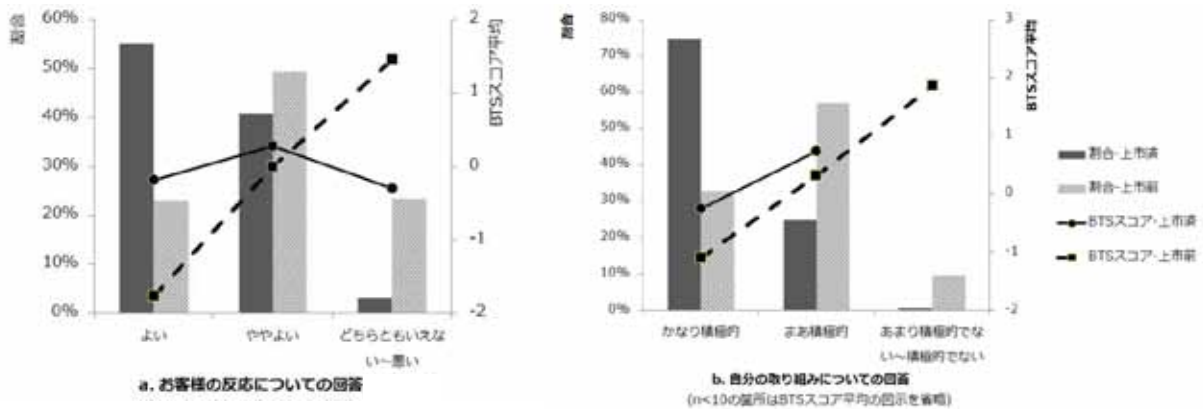


図2. 実験2の結果

**【設問】** 本報告では、調査に含まれていた次の4つの設問に注目する。

**現在販売中のある製品について**

- Q1. お客様の反応はいかがですか（「良い」から「悪い」までの5件法評定）
- Q2. あなた自身の現在の取り組みはいかがですか（「かなり積極的」から「積極的でない」までの4件法評定）

**来期に発売される、現在紹介活動中の新製品について**

- Q3. お客様の反応はいかがですか（「良い」から「悪い」までの5件法評定）
- Q4. あなた自身の現在の取り組みはいかがですか（「かなり積極的」から「積極的でない」までの4件法評定）

本実験では、これらの4つの設問の直後に、他の販売員の回答を予測する設問を追加した。たとえばQ1の直後には、Q1の5つの選択肢について、他の回答者がその選択肢を選ぶ割合を予測するように求めた。回答に際しては、0%から100%までの5%刻みの選択肢から一つを選択させた。

**【仮説】** これらの設問への回答における評価懸念バイアスについて、次のように予想される。

顧客の反応についての評価においては、上市前製品（Q3）については強い評価懸念バイアスが働くが、上市済製品（Q1）については働かないものと考えられる。上市済製品はすでに販売実績が明らかであるため、製品への顧客の反応についてどのように回答しても、自身のモチベーションの低さの現れとして解される心配はない。

いっぽう自分の取り組みについての評価においては、上市前製品（Q4）、上市済製品（Q2）ともに、強い評価懸念バイアスが働くものと考えられる。

**【結果と解釈】** Q1-Q4について、回答の分布、ならびに各回答に付与されたBTSスコアの平均を示す（図2）。

顧客の反応についての設問（図2a）では、上市前製品に対してはポジティブな回答におけるBTSスコアの平均が低くなった。いっぽう上市済製品については、いずれの回答でもBTSスコア平均に大差がなかった。

自分の取り組みについての設問（図2b）では、

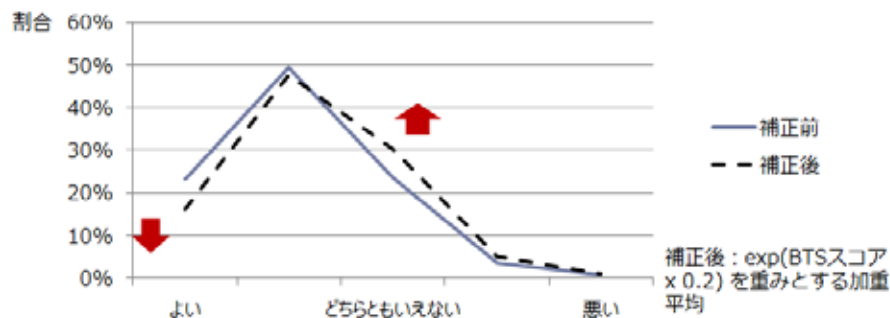


図3. 実験2における調査結果の補正（上市前製品への顧客の評価）

上市前製品、上市済製品ともに、ポジティブな回答における BTS スコア平均が低くなった。

これらの結果は、事前に予想された評価懸念バイアスの方向と整合している。上市前製品への顧客の反応についての評価、ならびに上市前・上市済製品に対する自分の取り組みについての評価において、評価懸念により回答がポジティブな方向に歪み、結果として、ポジティブな回答の真実性は低くなっているものと推測できる。BTS スコアは、評価懸念バイアスによる回答の真実性の低下を捉えているものと解釈できる。

なおこの観点からは、BTS によって集計結果を補正するという活用方法が考えられる。図 3 は、BTS スコアを重みとして回答を加重平均した結果を示している。BTS による補正によって、上市前製品への顧客の反応評価における「よい」率が減少し、「どちらともいえない」率が増大している。

## 実験 2： ネット調査パネルによる個人視聴率の予測

本実験では、消費者調査による行動意向の測定という文脈において、BTS の適用可能性を検討する。

**【背景】** 人々がある行動をとる意向を持っているかどうかを調べるという問題は、経済学における支払意思額測定や、世論調査・選挙調査など、さまざまな文脈で登場する。マーケティングの文脈でも、企業は人々がある消費者行動（たとえば新製品の購入）を行う意向があるかどうかに関心を持つ。

行動意向は必ずしも実際の行動を意味せず、その関係性の強さは状況によってさまざまである。たとえば投票行動は、事前の投票への意向だけではなく当日の天候にも影響されるだろうし、消費財の購買行動は、消費者の購買意向だけではなく、店頭への配架など流通上の要因にも強く影響されるだろう。しかし、仮に行動意向と実際の行動との関連性が弱いとしても、行動意向の理解それ自体が意義を持つ場合が少なくない。人々の投票意向は政治意識の指標としての意義を持ち、新製品への購入意向は製品評価の重要な指標となる。

本実験では、行動意向測定の事例として、TV 視聴者の番組視聴意向の測定という題材に注目し、BTS の適用による調査結果の補正を試みる。

行動意向の回答が一般にそうであるように、視聴意向の回答は実際の視聴行動と一致しない。そのずれは、回答の不正確さだけでなく、意向以外の諸要因によっても引き起こされているだろう。したがって、仮に完全に正直かつ真剣な回答が得られたとしても、測定された視聴意向が実際の視聴行動と完全に一致することは期待できない。しかし、BTS によって回答の真実性を把握できるならば、BTS による集計結果の補正を通じて、意向と行動とのずれが小さくなることが期待される。

なお、本実験では BTS のほかに、BTS のために調査票に追加する、他者の回答について予測する設問への回答分布そのものにも注目する。選挙調査の文脈では、自分の投票意向よりも他者の投票行動の予測のほうが実際の得票率と相関するという現象が報告されており、選挙結果予測の手法と

表 1. 実験 2 における聴取対象番組

	番組	分析対象 回答数
1	『おんな城主 直虎』（NHK総合 日曜20:00～ 主演 柴咲コウ）	779
2	『大貧乏』（フジテレビ 日曜21:00～ 主演 小雲）	815
3	『A LIFE 愛しき人』（TBS 日曜21:00～ 主演 木村拓哉）	815
4	『視覚探偵 日暮旅人』（日本テレビ 日曜22:00～ 主演 松坂桃李）	853
5	『突然ですが、明日結婚します』（フジテレビ 月曜21:00～ 主演 西内まりや）	2069
6	『嘘の戦争』（フジテレビ 火曜21:00～ 主演 草薙剛）	2069
7	『カルテット』（TBS 火曜22:00～ 主演 松たか子）	2069
8	『東京タラレバ娘』（日本テレビ 水曜22:00～ 主演 吉高由里子）	2069
9	『就活家族 きっと、うまくいく』（テレビ朝日 木曜21:00～ 主演 三浦友和）	2069
10	『嫌われる勇気』（フジテレビ 木曜22:00～ 主演 香里奈）	2069
11	『三匹のおっさん3 正義の味方、みたび!!』（テレビ東京 金曜20:00～ 主演 北大路欣也）	2069
12	『お母さん、娘をやめていいですか?』（NHK総合 金曜22:00～ 主演 波瑠）	2069
13	『下剋上受験』（TBS 金曜22:00～ 主演 阿部サダヲ）	2069
14	『精霊の守り人 悲しき破壊神』（NHK総合 土曜21:00～ 主演 綾瀬はるか）	2069
15	『スーパーサラリーマン左江内氏』（日本テレビ 土曜21:00～ 主演 堤真一）	2069

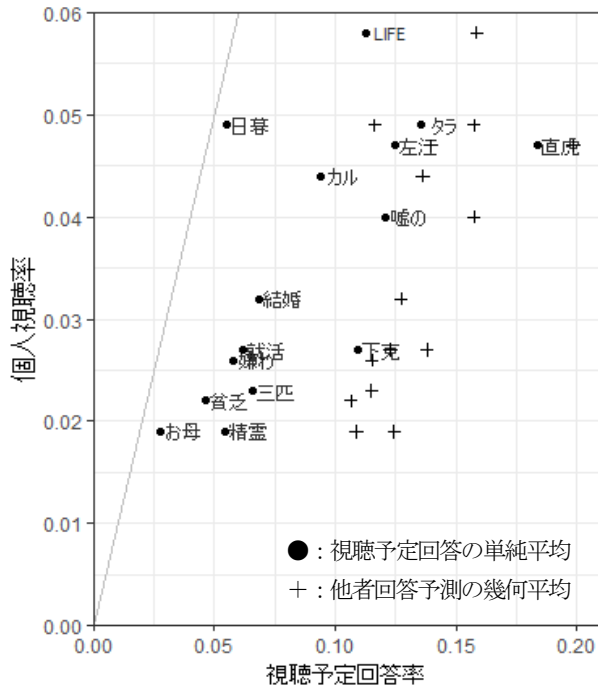


図4. 実験2の結果

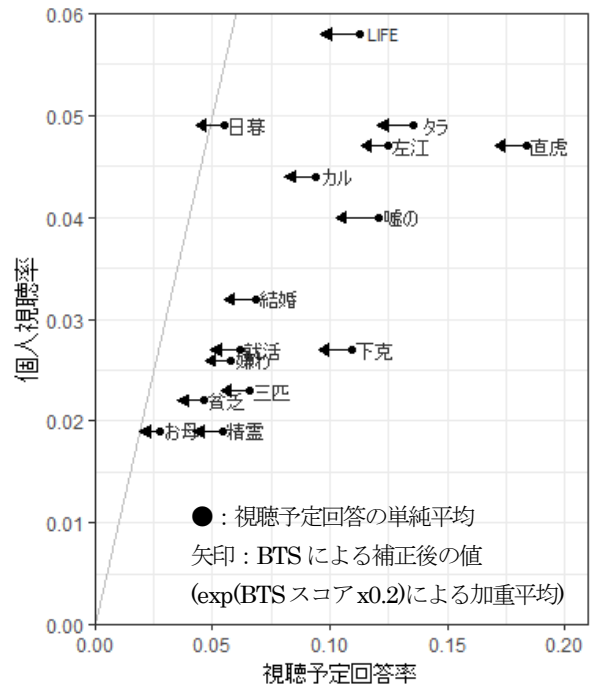


図5. 実験2の結果のBTSによる補正

して適用されている (vote expectation survey; Graefe, 2014).

**【方法】** 公募型ネットパネルから得た関東地方の15-64歳男女2,069名に対してネット調査を行った。調査は2017年1月21~23日に行った。

調査では、これから放映される連続TVドラマ15番組について聴取した。回答終了の時点ですでに放映が始まっていた回答は分析から除外する。番組名と分析対象回答数を表1に示す。

**【設問】** 主な聴取項目は、各番組について次の2問であった。

Q1. あなたは、今週放送される予定の次のテレビドラマを、リアルタイムで見る予定はありますか。  
(「リアルタイムで見る」「見ない」の2件法)

Q2. さきほど、あなたが今週放送される予定の次のテレビドラマを、リアルタイムで見るかどうか回答していただきましたが、この質問について、このアンケートに参加している他の人たちのうち、何%くらいの方が、「リアルタイムで見る」と回答すると思いますか。「リアルタイムで見る」と回答する人の割合を予想して、お答えください。  
(0%から100%までのスライダー)

**【外的基準】** 人々の実際の視聴行動を表す指標として、スイッチ・メディア・ラボ社のテレビ視聴分析システム SMART (<https://www.switch-m.com/>) が提供する個人視聴率を使用した。このシステムは、関東地方5,000人について個人レベルのリアルタイム視聴データを提供しており、ネットパネルに対する調査との比較が容易である。

**【結果】** いずれの番組でも、視聴予定回答率は実際の個人視聴率を上回った。また、他者の回答についての予測は、視聴予定回答率をさらに上回る値となった(図4)。すなわち、本研究では vote expectation survey の有効性は示されなかった。

BTS スコアによって視聴予定回答率を補正すると、実際の個人視聴率とのずれは小さくなった(図5)。この結果は、BTS スコアが回答の真実性を表すという想定と整合する。

回答に付与されたBTSスコアを観察すると、当該の連続テレビドラマを先週までに視聴している視聴者において、視聴予定回答率がきわめて高く、また予測スコア・情報スコアがともに低くなっていた(表2)。先週までに視聴した人は、自分の視聴確率を(おそらく実際よりも)高く見積もっており、BTSスコアはこれらの人々の回答が及ぼす影響を割り引く役割を果たしているものと思われる。



表 2. これまでの視聴と BTS スコアの関係

当該番組に対する これまでの視聴行動	割合	(視聴予定回答 率)/(実際の個人 視聴率)	予測スコ ア	情報スコ ア	BTSスコ ア
先週の放送をリアルタイムで見た	7%	25.5	-0.51	-0.35	<b>-0.86</b>
先週の放送を録画して後で見た	6%	5.0	-0.28	0.12	<b>-0.16</b>
先週の放送を録画したがまだ見ていない	3%	4.1	-0.25	0.13	<b>-0.12</b>
先週の放送はみていないが、それ以前の放 送は見たことがある	2%	6.5	-0.20	0.09	<b>-0.11</b>
見たことがない・知らない	82%	0.4	-0.15	0.23	0.08

(番組 4,5 を除く 13 番組の平均)

## 5. 考察

本研究では、マーケティング意思決定支援の文脈での消費者調査・従業員調査における、ベイジアン自白剤の適用可能性について検討した。

実験1では、従業員による製品評価における BTS スコアが、評価懸念バイアスが想定される設問において低くなることが示された。また実験2では、TV 視聴者における番組視聴意向の回答における BTS スコアが、過去経験に基づく強いバイアスを持つと想定される視聴者において低くなることが示された。以上の結果は、回答の真実性をあらわす指標としての BTS スコアの有用性を支持するものであるといえる。

従来の調査法研究は、調査主体が関心を持つなんらかの知識・態度を、調査回答者が調査参加という行為に先行して保持しており、調査主体の求めに応じてそれを開示する、という枠組みに沿って進められていた。この枠組みでは、調査という行為の本質的な相互作用性にはあまり注意が向けられなかった。これに対して、近年の調査法研究では、調査という行為を社会的相互作用とみなす観点から、調査の改善に寄与する新しい手法が生み出されている。その一つであるベイジアン自白剤は、回答品質の評価と改善という問題に対する、外的基準やパラ・データに依存しないアプローチとして、高い有用性を持っているものと思われる。

調査実務におけるベイジアン自白剤のさらなる活用のためには、以下に挙げる点について、さらなる実証的検討が必要である。

第一に、本研究のようにベイジアン自白剤を回答品質の評価手法として捉える際には、調査結果の補正にあたって BTS スコアをどの程度まで考慮するかという点についての基準が必要となるだろ

う。本研究では先行研究を参考に、BTS スコアを指数変換した値を重みとして回答を加重平均するという方法を採用したが、この方法に実証的な根拠はない。この問題は、回答の真実性と BTS スコアとの定量的関係という問題であるといえる。本研究のようなフィールド実験だけでなく、真の回答分布が既知であるような実験的状況を用いた実証研究の蓄積が必要であろう。

第二に、ベイジアン自白剤を回答品質の改善手法として捉える方向についても、実証研究がまだまだ不足している。とくに、BTS スコアをインセンティブと連動させることによる回答品質の改善は、意図されているようなインセンティブ整合的メカニズムの働きによるものなのか、それとも「BTS スコアが正直さを表す」という教示の効果に過ぎないのかという点について、従来の研究には曖昧さが残されている (Kuncel, Borneman, Kiger, 2012)。心理学研究においては、態度測定において「あなたの本当の気持ちを測定する装置があります」という(偽りの)教示を行うことで回答が正直な方向へと変動することが知られている (bogus pipeline 効果 Cf. Roese & Jamieson(1993))。仮にベイジアン自白剤の効果が一種の bogus pipeline 効果に過ぎないのであれば、その適用可能性には限界があるものと思われる。

第三に、ベイジアン自白剤における回答の心理過程についての検討、いいかえればベイジアン自白剤の個人内での作用機序の解明が求められる。回答者は、他者回答の予測という設問に対してどのような方略を使用して回答しているのか。また BTS スコアが報酬と連動する状況下で、回答者は BTS スコアをどのような値として理解しているのか。BTS スコアが低くなるのは、回答にあたって

の誠実性が欠けているからか、それとも回答にあたっての十分な知識が欠けているからか。これらの点についての実証研究は、調査においてベイジアン自白剤を適用しやすい場面とそうでない場面について理解する際に、重要な示唆を提供するものと思われる。また、BTS スコアを回答時間など別の指標と併用する際の方法についても示唆を提供するだろう。

最後に、ベイジアン自白剤を予測メカニズム一般のなかに位置づける理論的枠組みの整備が必要である。巨視的観点から見ると、多くの調査は、なんらかの事象を予測するために構築された、集合知に基づく予測メカニズムの一形態であると捉えることができる。Chen & Pennock (2010)はメカニズム・デザイン論の観点から予測メカニズムを類型化し、ベイジアン自白剤をピア予測システムの一つとして位置付け、予測市場などの他のメカニズムと比較している。また Jurca & Falting (2008)は、同じくメカニズム・デザイン論の観点からベイジアン自白剤を改訂し、真実申告がベイジアン・ナッシュ均衡解となるという性質を放棄するかわりに、調査終了を待たずに報酬を決定される手法を提案している。調査を集合知メカニズムの一つとして捉えるこうした観点は、従来の調査法研究においては見失われがちであった視点であり、今後の調査手法開発に新しい可能性をもたらすだろう。

(株式会社インサイト・ファクトリー)

注) 本報告の一部は下記の発表に基づいている：  
小野滋(2015)「消費者調査における真実申告誘発メカニズムの活用」, 日本行動計量学会第43回大会, 2015.09.

小野滋(2017)「消費者調査における真実申告メカニズムの活用～ベイジアン自白剤によるテレビ視聴行動予測～」, 日本行動計量学会第45回大会, 2017.08.

## 参考文献

- Becker, G.M., DeGroot, M.H., Marschak, J. (1964) Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3), 226-232.
- Chen, Y., Pennock, D.M. (2010) Designing markets for prediction. *AI Magazine*, 31(4).
- Ding, M. (2007) An incentive-aligned mechanism for conjoint analysis. *Journal of Marketing Research*, 44(2), 214-223.

- Frank, M.R., Cebrian, M., Pickard, G., Rahwan, I. (2017) Validating Bayesian truth serum in large-scale online human experiments. *PRoS ONE*, 12(5), e0177385.
- Graefe, A. (2014) Accuracy of vote expectation surveys in forecasting elections. *Public Opinion Quarterly*, 78, 204-232.
- Howie, P.J., Wang, Y., Tsai, J. (2011) Predicting new product adoption using Bayesian truth serum. *Journal of Medical Marketing*, 11(1), 6-16.
- John, L.K., Lowenstein, G., Prelec, D. (2012) Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524-532.
- Jurca, R., Falting, B. (2008) Incentives for expressing opinions in online polls. *EC '08: Proceedings of the 9th ACM conference on Electronic commerce*. 119-128.
- Kuncel, N.R., Borneman, M., Kiger, T. (2012) Innovative item response process and Bayesian faking detection methods: More questions than answers. in Ziegler, M., Maccann, C., Roberts, R.D. (eds.) *New perspectives on faking in personality assessment*. Oxford University Press.
- Miller, S.R., Brailey, B.P., Kirlik, A. (2014) Exploring the utility of Bayesian truth serum for assessing design knowledge. *Human-Computer Interaction*. 29(5-6), 487-515.
- Prelec, D. (2004) A Bayesian truth serum for subjective data. *Science*, 306(5695), 462-466.
- Roese, N.J., Jamieson, D.W. (1993) Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin*, 114(2), 363-375.
- Show, A.D., Horton, J.J., Chen, D.L. (2011) Designing incentives for inexpert human raters. *CSCW'11: Proceedings of the 2011 ACM Conference on Computer Supported Cooperative Work*. 275-284.
- Weaver, R., Prelec, D. (2013) Creating truth-telling incentives with the Bayesian truth serum. *Journal of Marketing Research*, 50(3), 289-302.