

令和 6 年 6 月 4 日現在

機関番号：14401

研究種目：研究活動スタート支援

研究期間：2020～2023

課題番号：20K22268

研究課題名（和文）確信度に基づいた嘘検知の認知プロセスの解明

研究課題名（英文）An examination of the cognitive process of deception detection based on confidence

研究代表者

大工 泰裕 (Daiku, Yasuhiro)

大阪大学・大学院人間科学研究科・招へい研究員

研究者番号：30880322

交付決定額（研究期間全体）：（直接経費） 2,200,000円

研究成果の概要（和文）：本研究では、真偽判断における確信度が正答率を予測するというSmith and Leach (2019)の結果を、ベイズ統計モデリングを用いて再検討した。1. 確信度にかかわらず正答率が一定であるモデル、2. 確信度の高さが正答率の高さを予測するロジスティック回帰モデル、3. 2のモデルの偏回帰係数がある分割点において変化するという3種類のモデルを比較した結果、3のモデルの当てはまりが最も良かったが、分割点や偏回帰係数の95%HDIが非常に大きいことから、確信度が正答率に影響を与えているとは言えなかった。このことから、真偽判断の際に根拠に基づいた判断を行っている可能性は低いことが示唆された。

研究成果の学術的意義や社会的意義

嘘検知を扱った先行研究では、古典的な研究手法によって、要因操作と正答率の因果関係という入力と出力の関係のみに焦点が当てられてきており、真偽判断の認知プロセスそのものは研究対象になってこなかった。本研究は、近年指摘され始めた確信度と正答率の関係を、ベイズ統計モデリングという新しい手法で精緻に検討することによって、真偽判断の認知プロセスそのものに焦点を当てた。確信度と正答率の確固たる関係性は見られなかったものの、嘘検知の認知プロセスの検討にベイズ統計モデリングという新たな方法を持ち込み、嘘研究との親和性の高さを示した点で学術的意義がある。

研究成果の概要（英文）：In this study, I revisited Smith and Leach's (2019) result that confidence in veracity judgments predicts accuracy using Bayesian statistical modeling with the following three models: 1. a model in which accuracy is constant regardless of confidence; 2. a logistic regression model in which higher confidence predicts higher accuracy; and 3. a segmented logistic regression model in which higher confidence predicts higher accuracy but the coefficient changes at a breakpoint. Model 3 was the best fit, but the very wide 95% HDI of the breakpoint and the coefficient did not indicate that confidence predicted accuracy. This suggested that it was unlikely that the judges were making evidence-based veracity judgments.

研究分野：社会心理学

キーワード：嘘 欺瞞 虚偽検出 真偽判断 統計モデリング

1. 研究開始当初の背景

嘘検知は、心理学の中でも古典的なテーマの1つである。そのほとんどにおいて、どうすれば他者の行動から発言の真偽を見分けることができるかということに研究の焦点が置かれてきた (Levine & McCornack, 2014)。それゆえ、嘘検知の研究手法もこの目的に即したものとなっており、送り手が話している様子(動画・対面等)を見て、受け手がその真偽を判断するという実験室実験が長年用いられてきた。この手法では、1人の受け手が複数の送り手に関する真偽判断を行い、各試行において受け手の真偽判断が送り手の真値とどの程度一致していたかを表す正答率が従属変数とされる。

Bond & DePaulo (2006) は、このような手法を用いた 206 もの嘘検知に関わる研究をメタ分析し、嘘検知の正答率はチャンスレベルをわずかに上回る 54%であることを発見した。人間に嘘検知の能力が全く備わっていないのであれば、正答率は 50%となるはずだが、それをわずかではあれ上回るこの現象は、中程度の効果量 ($d=0.40$)を持って多くの研究で観察される (Levine, 2020)。しかし、この現象を生み出すプロセスは長らく解明されてこなかった。

近年、この現象の解明につながるヒントとして、Smith & Leach (2019) は受け手の判断の確信度が正答率に影響することを発見した。彼らは送り手の発言の真偽を受け手に判断させる際、どの程度その判断に自信があるかを 0-100%で同時に評定させた。確信度によって「70%未満」「70-89%」「90%以上」の3つに試行を分割した結果、確信度が「70-89%」と「90%以上」の試行では、正答率が高くなっていった一方、確信度が「70%未満」のときは正答率が低くなっていった。

2. 研究の目的

Smith & Leach (2019) の研究結果は、確信度の高い判断に限れば、人間は他者の行動から発言の真偽を正しく判断できる傾向があることを示している。これは、確信度が高い場合、受け手が送り手の真偽を示す何らかの根拠を送り手の行動の中に発見し、それに基づいた真偽判断をした結果、正答率が向上するからだと推測される。一方で、確信度が低い場合、根拠を見つけられずに真偽判断をした結果として正答率が低下すると考えられる。つまり、従来の嘘検知研究における真偽判断の正答率は、明確な根拠に基づいて行われる判断と根拠に基づかない判断という2種類のプロセスが合わさって生成されていることが示唆される。

しかし、Smith & Leach (2019) では確信度を恣意的に分割して検証しているという問題があった。真偽判断に2種類のプロセスが存在していたとしても、それが確信度のどこを閾値として分割すれば検証できるかは定かではない。そこで、申請者は統計モデリングを用いてこの閾値を客観的に推定し、50%よりわずかに高い真偽判断の正答率が2種類の真偽判断プロセスが合わさった結果として説明可能かをより精緻に検討することを目的とした。

具体的には、以下のような3種類のロジスティック回帰モデルにデータを当てはめ、モデル比較によって最良なモデルを選出した。

1. $\text{logit}(\pi_{ij}) = \beta_{0j}$
2. $\text{logit}(\pi_{ij}) = \beta_{0j} + \beta_1 \text{Confidence}_{ij}$
3.
$$\begin{cases} \text{logit}(\pi_{ij}) = \beta_{0j} + \beta_1 \text{Confidence}_{ij}, & \text{if } \text{Confidence}_{ij} \leq \alpha \\ \text{logit}(\pi_{ij}) = \beta_{0j} + \beta_2 \text{Confidence}_{ij}, & \text{otherwise} \end{cases}$$

ここで、

$$y_{ij} \sim \text{Bernoulli}(\pi_{ij})$$
$$\beta_{0j} = \beta_0 + u_j, u_j \sim N(0, \sigma_u^2)$$

であり、 y は正答率、 i は各試行、 j は各動画、 α は 2 種類の真偽判断のプロセスが切り替わる分割点を表す。動画によるランダム切片のみを仮定したのは、先行研究で、真偽判断の受け手の分散は小さいが送り手の分散が大きいことが示されているためであった (Bond & Depaulo, 2008)。

各モデルについての説明を述べる。1 は正答率が確信度に関係なく一定であるモデルである。2 は確信度が上昇するとともに正答率が一定割合で向上するロジスティック回帰モデルである。3 は分割点を境にして偏回帰係数が異なるロジスティックモデルであり、Smith & Leach (2019) の知見が正しければ、このモデルが最も適合度の高いモデルとなり、 β_1 の 95%HDI が 0 を含み、 β_2 の 95%HDI が 0 を含まない結果になると予測される。

3 . 研究の方法

参加者

2022 年 8 月、クラウドソーシングサービスサイト Prolific を利用し、米国の英語話者 200 名を対象に Web 実験を実施した。実験の所要時間は約 40 分で、参加者は報酬として 6 ドルを得た。

刺激

嘘研究用の実験刺激をまとめている Miami university deception detection database (Lloyd et al., 2019) から、16 の動画を選出して実験刺激とした。選定基準は、動画の人物の性別 (男性・女性)、動画の人物の人種 (黒人・白人)、動画の感情価 (Positive・Negative)、動画の人物が話している内容の真偽 (Truth・Lie) の各カテゴリーがそれぞれ 1 つずつ含まれるように抽出した。なお、それぞれのカテゴリーの中で、正答率ができる限り 50% に近いものを抽出した。その結果、BF006_1PT、BF020_4PL、BF023_3NT、BF024_2NL、BM006_3NT、BM018_4PL、BM020_1PT、BM027_2NL、WF009_4PL、WF011_1PT、WF018_2NL、WF025_3NT、WM002_2NL、WM006_1PT、WM007_3NT、WM008_4PL の 16 の動画を最終的に刺激として用いた。また、BF001_1PT を実験のインストラクション用の動画として用いた。

手続き

実験は Qualtrics を用いて構築した。参加者は Web から実験ページにアクセスし、まず実験に関する説明とデータの取扱いについての説明を受け、同意を行った場合のみ実験に参加した。質問への回答の仕方に関する簡単なインストラクションを受けた後、上述した 16 の動画を再生し、それぞれの動画に対して「動画上の人物が話している内容が真実か嘘か」を 2 件法で、「その判断にどれくらいの自信があるか」を 0 から 100 までの 10 刻みの 11 件法で回答した。16 の動画それぞれについて視聴と回答を終えた後、性別・年齢を回答して実験は終了した。

なお、参加者が動画を視聴せずに回答することを避けるため、動画と回答ページを別ページとし、動画の再生秒数が過ぎるまでは回答ページに移行できないように設定した。また、16 の動画の提示順序はランダムマイズされていた。

分析対象者

Prolific 上の回答終了データと突合できなかった 2 名、調査回答終了までの回答時間が平均回答時間 + 2SD (5606 秒) 以上であった参加者 9 名の合計 11 名を分析から除外した。最終的な分析対象者は 189 名 (男性 93 名、女性 96 名、平均年齢 33.44 ± 11.28 歳)、分析対象試行は 3024 試

行であった。

4. 研究成果 分析

Table1 に動画ごとの正答率 (Accuracy), 真と回答した割合 (Truth Bias), 確信度 (Confidence) を示した。Lloyd et al. (2019) から正答率が 50% に近い動画を選出したが、全体的に正答率が 50% から大きく乖離する結果となった。

Table1 動画ごとの要約統計量

Video	Accuracy	Truth Bias	Confidence(mean)	Confidence(SD)
BF006_1PT	0.61	0.61	65.4	22.0
BF020_4PL	0.47	0.53	64.1	19.9
BF023_3NT	0.66	0.66	67.4	21.9
BF024_2NL	0.41	0.59	63.9	22.4
BM006_3NT	0.61	0.61	61.0	20.5
BM018_4PL	0.64	0.37	60.7	21.7
BM020_1PT	0.66	0.66	64.6	20.8
BM027_2NL	0.47	0.53	63.9	20.9
WF009_4PL	0.56	0.44	65.4	21.6
WF011_1PT	0.64	0.64	69.3	20.7
WF018_2NL	0.42	0.58	62.7	21.3
WF025_3NT	0.64	0.64	65.4	19.7
WM002_2NL	0.41	0.59	60.8	20.7
WM006_1PT	0.60	0.60	62.7	22.9
WM007_3NT	0.55	0.55	61.5	20.1
WM008_4PL	0.68	0.32	61.5	22.9

次に、PyMC を用いて各モデルの比較を行った。事前分布としては、以下を用いた。

$$\sigma_u \sim \text{Half-Cauchy}(5)$$

$$\beta_1 \sim N(0, 10^2)$$

$$\beta_2 \sim N(0, 10^2)$$

$$\alpha \sim U(0, 1)$$

バーンインは 2000、サンプリング回数は 4000、チェーン数は 4 で推定を行った。いずれのモデルも Rhat が 1.1 以下になることとトレースプロットによって収束を確認した。

各パラメータの推定結果およびモデルの WAIC を Table2 に示す。Model 3 の WAIC が最も低くなっており、Model 3 がデータに最も適合するモデルであることが示された。しかし Model3 は β_2 および α の 95% HDI が非常に大きくなっていった。

(データ処理に誤りがあったため再分析を実施した結果、最終年度の実績報告書時点と当てはまりの良いモデルが異なっています)

Table2 各モデルのパラメータの推定値およびモデルの WAIC

	mean	sd	95% HDI	WAIC
Model 1				-2031.94
σ_u	0.48	0.11	[0.30–0.69]	
Model2				-2032.29
β_1	0.003	0.001	[0.0005–0.0054]	
σ_u	0.41	0.10	[0.24–0.60]	
Model3				-2032.32
β_1	0.003	0.001	[0.0005–0.0054]	
β_2	0.08	10.08	[-19–20]	
σ_u	0.41	0.10	[0.24–0.59]	
α	0.50	0.29	[0.0004–0.95]	

考察

以上より、Smith & Leach (2019) の知見を仮定した Model 3 が最もデータに当てはまりが良かったが、分割点の推定には成功しておらず、 β_2 も 0 よりも大きいとは言えなかった。 β_1 についても 0 に限りなく近い値となっていたことから、確信度は正答率に影響を与えていないと結論づけるのが妥当であろう。以上の結果から、人間は真偽判断を行うときに明確な根拠に基づいて判断しているとは言い難いことがわかった。

この結果は嘘検知の実験パラダイム自体から来ている現象であるとも考えられる。嘘検知の実験では、本研究のように見ず知らずの人物の言動が嘘かどうかを判断させられることが一般的である。人間が人物の普段の行動との乖離や、物理的な証拠に基づいて嘘を見破っていることを考えると、この実験パラダイムは不自然である (Levine, 2018)。そうした、生態学的妥当性の低い実験パラダイム自体が、根拠に基づいた真偽判断を難しくしているとも考えられるため、より生態学的妥当性の高い状況を仮定した実験パラダイムが今後必要とされる。

引用文献

- Bond, C. F., & DePaulo, B. M. (2006). Accuracy of Deception Judgments. *Personality and Social Psychology Review, 10*(3), 214–234.
- Bond, C. F., & DePaulo, B. M. (2008). Individual Differences in Judging Deception: Accuracy and Bias. *Psychological Bulletin, 134*(4), 477–492.
- Levine, T. R. (2018). Ecological Validity and Deception Detection Research Design. *Communication Methods and Measures, 12*(1), 45–54.
- Levine, T. R., (2020). *Duped: Truth-default theory and the social science of lying and deception*. Tuscaloosa, AL: University of Alabama Press.
- Levine, T. R., & McCornack, S. A. (2014). Theorizing About Deception. *Journal of Language and Social Psychology, 33*(4), 431–440.
- Lloyd, E. P., Deska, J. C., Hugenberg, K., McConnell, A. R., Humphrey, B. T., & Kunstman, J. W. (2019). Miami University deception detection database. *Behavior Research Methods, 51*, 429–439.
- Smith, A. M., & Leach, A. M. (2019). Confidence Can Be Used to Discriminate Between Accurate and Inaccurate Lie Decisions. *Perspectives on Psychological Science, 14*(6), 1062–1071.

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件／うち国際共著 2件／うちオープンアクセス 3件）

1. 著者名 Murai Jun' ichiro, Daiku Yasuhiro	4. 巻 72
2. 論文標題 Diary Study on Japanese University Students' Suspicions of Being Deceived in Everyday Life	5. 発行年 2021年
3. 雑誌名 Communication Studies	6. 最初と最後の頁 807～818
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/10510974.2021.1975147	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Murai Jun' ichiro, Daiku Yasuhiro	4. 巻 2
2. 論文標題 Deception in everyday life of Japanese young adults: a reanalysis	5. 発行年 2021年
3. 雑誌名 Discover Psychology	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s44202-021-00018-y	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Ueno Daisuke, Daiku Yasuhiro, Eguchi Yoko, Iwata Minako, Amano Shoka, Ayani Nobutaka, Nakamura Kaeko, Kato Yuka, Matsuoka Teruyuki, Narumoto Jin	4. 巻 12
2. 論文標題 Mild Cognitive Decline Is a Risk Factor for Scam Vulnerability in Older Adults	5. 発行年 2021年
3. 雑誌名 Frontiers in Psychiatry	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.3389/fpsy.2021.685451	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Yasuhiro Daiku, Kim B. Serota, Timothy R. Levine	4. 巻 16
2. 論文標題 A few prolific liars in Japan: Replication and the effects of Dark Triad personality traits	5. 発行年 2021年
3. 雑誌名 PLOS ONE	6. 最初と最後の頁 e0249815
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pone.0249815	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1. 著者名 Timothy R. Levine, Yasuhiro Daiku, Jaume Masip	4. 巻 -
2. 論文標題 The numbers of senders and total judgments matter more than sample size in deception detection experiments	5. 発行年 2021年
3. 雑誌名 Perspectives on Psychological Science	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕 計1件 (うち招待講演 0件 / うち国際学会 0件)

1. 発表者名 大工泰裕
2. 発表標題 A few prolific liarsの再現と方法論的課題
3. 学会等名 日本心理学会第85回大会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------