

他者の目効果が不正及び自白行為に与える影響に関する検証*

中川大地¹ 松田夕季² 笹田実里³

要約

本研究は、デジタル上における他者の目効果が、不正及び自白行為に与える影響を検証する。近年、企業における不正が深刻な問題となっており、企業は適切な対策を講じることが求められている。さらにデジタル上においても、人々は他者の目効果の影響により、対人間に比べ、対機械で3倍不正行為を行うことが示された。また、AI エージェントはデジタル上で人々の行動に影響を与えることが明らかになった。そこで本研究では、AI エージェントを他者の目として捉え、AI エージェントが参加者を見ている、参加者を見ていない、いないの3群間で、不正しにくさ及び自白しやすさについて検証した。その結果、「見ていない」群において最も不正及び自白をしにくいことが確認され、他者の目効果の影響を考察した。本研究は、AI エージェントによる他者の目効果が、不正行為の抑制及び自白行為の促進に与える影響について一定の示唆を与える。

JEL 分類番号 : D90, D91

キーワード : 他者の目効果, 不正, 自白, AI エージェント

¹ 中川大地 同志社大学商学部 cgfg0407@mail3.doshisha.ac.jp

² 松田夕季 同志社大学商学部 cgfg0314@mail3.doshisha.ac.jp

³ 笹田実里 同志社大学商学部 cgfg0553@mail3.doshisha.ac.jp

*本論文に関して、開示すべき利益相反関連事項はない。

1. イントロダクション

1.1. 不正の現状

近年、企業の不祥事が相次いで発覚している。KPMG (2022) が実施した日本企業の不正に関する実態調査によると、上場企業 578 社中 137 社 (約 24%) が、直近 3 年間で不正が発生したと回答している。また、デジタル化に伴い、約 93% の企業が IT データを活用した不正検知に関心があると回答しており、デジタル技術を活用した不正対策は必要不可欠だ。

さらに、不正の開示に関して、東京証券取引所の規定では、投資者の投資判断に大きな影響を与える重要な事項が発生した場合、企業は速やかにその内容を開示しなければならないと定められている [稲葉 2016]。そのため、企業において不正の開示や、自白を促進する必要性は益々高まっている。

1.2. 他者の目効果

Jantsje et al. (2020) によると、能動的なアバター (参加者を見ている) では、受動的なアバター (参加者を見ていない) よりも不正行為が少なく、アバターなしのコントロール条件よりも不正行為が少ないことが示された。これは、他者に見られていると感じることで利他的な行動が促進される「他者の目効果」 [Haley & Fessler 2005] が影響している。

また、Cohn et al. (2021) は、デジタル上の不正に関して、人々は人間ではなく機械と相互作用する場合、約 3 倍多く不正行為を行うことを示した。また、不正を行う機会に直面した際、不正行為に及ぶ個人は機械との対話を好む傾向があることも分かった。このように不正をする際に対人間より対機械の方が好まれる背景には、他者からどう思われるかという社会的イメージを気にする傾向や心理バイアス (Social image concern) が影響している。

1.3. AI エージェント

梁・山田・寺田 (2015) によると、人々は推薦商品に対して元々もっている購買意欲に関わらず、商品推薦エージェントに推薦をされた場合に購買意欲が変化するという。また、津村・山田 (2021) の研究では、擬人化エージェントの外見が人間であり関連性が高い自己開示を行うことで共感的反応を促進できることが示唆された。以上から、AI エージェントの外見や行動は、人々の特定の行動を促進・抑制するための有効な手段となる可能性がある。

1.4. 本研究の目的・仮説

本研究では、デジタル上において AI エージェントによる他者の目効果が、人間の不正行為の抑制及び自白行為の促進にどのような影響を与えるのかを実験的に検証する。我々は、他者の目効果が利他的な行動を促進することに着目し、参加者に他者の視線を感じさせることで、不正行為を抑制できると考えた。他方、人々は社会的イメージを気にすることから、参加者に他者の視線を感じさせないことで、自白が促進されると考えた。これらを踏まえ、以下の仮説を設定した。

仮説 1 : AI エージェントが存在し、かつ参加者を見ていると不正をしにくい。

仮説 2 : AI エージェントが存在し、かつ参加者を見ないと自白しやすい。

2. サーベイ実験

2.1. 実験期間・参加者数

実験実施日は 2024 年 8 月 3 日であり、参加者は 1120 名で、不要な回答を除いた最終的な有効回答数は 10 代から 70 代の 445 名（女性 143 名、男性 299 名、その他 3 名）である。なお、不要な回答として、実験の内容をよく理解していないと見受けられるものや、回答不備があるものに加えて、出た目を正しく入力しなくても良いことに気がついていなかったと回答したものも除く[鶴田・犬飼 2019]。

2.2. 手続き

本実験は web 調査サイト GoogleForm を用いて作成し、Yahoo!クラウドソーシングにて配布を行った。初めに、参加者を AI エージェントが参加者を見ている群 (6.A 参照)、AI エージェントが参加者を見ない (目を瞑っている) 群 (6.B 参照)、AI エージェントがいない群の 3 群 (以下、「見ている」、「見えない」、「いない」群) に分け、それぞれに同様のタスク・質問に取り組んでもらう。「いない」群は統制群の役割を果たす。不正行動を測定するため、6 面サイコロを用いたタスクを実施した[Fischbacher and Föllmi-Heusi 2013]。

まず参加者に、サイコロの目が大きいほど報酬が高まる関係を示す報酬表 (6.C 参照) を提示し、デジタル上でサイコロを振らせ、その結果を入力させる (1 回目)。この際、「サイコロの出た目は参加者にしか確認できないこと」、「報酬ができるだけ高くなるような意思決定を行うこと」を伝える。次に、参加者が獲得する報酬が確定したことを伝え、実際に出たサイコロの目を再度入力させる (2 回目)。なお、報酬表の提示からサイコロの目の入力 (1 回目・2 回目) までの間、各群に対応する AI エージェントの画像は常に参加者に表示される。最後に個人属性を集計し、実験を終了する。

2.3. 測定

1 回目に報告されたデータにより参加者の不正度合い、2 回目に報告されたデータにより参加者の自白度合いを測定する。嘘の報告があったかどうかは、1 回目・2 回目で参加者が報告したサイコロの目のそれぞれの割合が、一様分布から乖離しているか否かで判別する。

まず、適応度検定により、不正及び自白度合いが、各群ごとに一様分布と乖離しているかを調べる。次に、フィッシャーの正確確率検定、多重比較により各群間で不正及び自白度合いに有意な差があるかを調べる。

3. 結果

3. 1. 記述統計

各群で参加者が報告したサイコロの目の有効回答数と、各群で参加者が報告したサイコロの目の有効回答数、平均値は、表1のようになった。

「表1 各群で報告されたサイコロの目の有効回答数」

サイコロの目	見ている群		見していない群		いない群	
	1回目	2回目	1回目	2回目	1回目	2回目
1	19	23	18	19	20	23
2	12	13	29	32	18	20
3	19	18	29	31	20	22
4	27	27	38	40	21	24
5	29	29	24	22	21	19
6	26	22	32	26	43	35
平均値	3.856	3.697	3.688	3.541	3.937	3.706

3. 2. 不正度合い

適応度検定の結果、「見ている」群で、統計量=9.4545、自由度=5、p値=0.09225、「見えない」群で、統計量=8.8947、自由度=5、p値=0.1133、「いない」群で、統計量=18.748、自由度=5、p値=0.002141となり、「いない」群において5%水準で、「見ている」群において10%水準で有意差が見られ、一様分布に従わないことが確認された。「見えない」群では有意差は見られず、一様分布に従うことが確認された。次に、フィッシャーの正確確率検定により、サイコロの目を6と報告した参加者の割合において、p値=0.04445となり、5%水準で3群間の不正度合いに有意差があることが確認された。最後に、多重比較を行った結果、「見ている」群と「見えない」群間は、p値=0.8835となり、有意差は見られなかった。

「見ている」群と「いない」群間は、p値=0.05214となり、10%水準で有意差が見られた。

「見えない」群と「いない」群間は、p値=0.02385となり、5%水準で有意差が見られた。

その上で、Holm補正を行うと、各群間で有意差は見られなかった。

3. 3. 自白度合い

適応度検定の結果、「見ている」群で、統計量=7.8182、自由度=5、p値=0.1665、「見えない」群で、統計量=10.018、自由度=5、p値=0.07474、「いない」群で、統計量=7、自由度=5、p値=0.2206となり、「見えない」群において、10%水準で有意差が見られ、一様分布に従わないことが確認された。「見ている」群、「いない」群では有意差は見られず、一様分布に従うことが確認された。次に、フィッシャーの正確確率検定により、サイコロの目を6と報告した参加者の割合において、p値=0.09801となり、10%水準で3群間の自白度合いに有意差があることが確認された。最後に、多重比較を行った結果、「見ている」群と「見えない」群間は、p値=0.7536となり、有意差は見られなかった。「見ている」群と「いない」群間は、p値=0.1364となり、有意差は見られなかった。「見えない」群と「いない」群間は、p値=0.04561となり、5%水準で有意差が見られた。その上で、Holm補正を行うと、各群間で有意差は見られなかった。

4. 考察

不正度合いに関して、「いない」群で最も不正が行われた。これは予想通りの結果である。しかし他方、「見ている」群と「見えない」群間において、多重比較では有意差は出な

ったものの、適応度検定の結果より、「見ている」群の方が、「見ている」群より不正が行われたと確認された。このことから、「見ている」群の方が「見ている」群よりも不正しにくいことが示される。これは我々の予想に反する帰結である。

自白度合いに関して、3群間において、多重比較では有意差は出なかったものの、適応度検定の結果より、「いない」群、「見ている」群、「見ている」群の順で自白が行われたと確認された。このことから、「見ている」群の方が「見ている」群よりも自白しやすいことが示される。これも我々の予想に反する帰結である。

以上より、我々の仮説は棄却され、不正しにくさ・自白しやすさともに、予想とは逆向きの結果が得られた。そこで、以下の2点について考察する。

考察1：AI エージェントが参加者を見ている方が、他者の目効果が弱い。

考察2：AI エージェントが参加者を見ている方が、他者の目効果が強い。

一つ目に関して、「見ている」群として用いたAI エージェントの画像は、少し口角を上げている表情であった。葛西・山本・倉本・辻野（2014）の研究により、対面する相手笑顔である場合、他者からの否定的な評価を恐れる気持ちが減るため、対人不安が減ることが明らかになっている。そのため、「見ている」群において、AI エージェントの表情によって、他者から受ける社会的イメージ（Social Image Concern）を気にしにくくなり、不正及び自白をしやすくさせたと考えられる。

二つ目に関して、「見ている」群として用いたAI エージェントの画像は、参加者の目の前にいながら目を瞑り、参加者の方を向いていないという不自然な状況であった。この違和感が参加者に過度な存在感を与え、目の前に人がいるという意識をより強く感じさせたため、不正及び自白をしにくくさせたと考えられる。

5. 研究の限界と今後の展望

本実験では、嘘行動の測定を全てデジタル上で実施した。デジタル上でのサイコロを振る体験では、参加者に十分な当事者意識を持たせられなかった可能性がある。また、報酬額についても、サイコロの目の報告に関わらず一定額しか支払われなかったため、参加者は短絡的に設問に回答してしまった可能性がある。また、岡田（2021）によると、デジタル上では見られているという実感が欠如し、集中力が低下することが指摘されている。そのため、AI エージェントに人格や表情を追加し、より人間らしい要素を持たせることや、動画形式を用いることで、参加者に他者の存在感を強く感じさせることができるだろう。

本研究で、不正のしにくさに関しては、「見ている」群で最も高い傾向が見られたため、デジタル上において「見ている」群のように違和感を覚えさせる現象を意図的に作り出すことで、人々の注意を引き、行動に影響を与えることができるだろう。一方、自白のしやす

さに関しては、「いない」群及び「見ている」群で高い傾向が見られたことから、不正抑制と自白促進には異なるアプローチが必要であることが明らかになった。今後は、デジタル上での行動制御や監視の効果を高める新たなアプローチを探求したい。

6. 付録

A. 「見ている」群で用いた画像, B. 「見ていない」群で用いた画像, C. 報酬表

A



B



C

サイコロの目	1	2	3	4	5	6
報酬	100	200	300	400	500	600

7. 引用文献

Alain Cohn, Tobias Gesche and Michel André Maréchal, 2021. Honesty in the Digital Age. *Management Science* Vol. 68, No. 2.

稲葉喜子, 2016. 不正会計を行う経営者の動機. *早稲田商学*第 446 号, 419-435 719-735.

Jantsje M. Mol, Eline C. M. van der Heijden and Jan J. M. Potters, 2020. (Not) alone in the world: Cheating in the presence of a virtual observer. *Experimental Economics*.

葛西響子, 山本景子, 倉本到, 辻野嘉宏, 2014. コウテイカボチャ: 聴衆に肯定的な反応を重畳する発表時緊張感緩和手法. *情報処理学会研究報告*, Vol.2014-HCI-160, No.8.

Kevin J. Haley and Daniel M. T. Fessler, 2005. Nobody's watching? Subtle cues affect generosity in. *Evolution and Human Behavior* 26(3), 245-256.

KPMG, 2022. *Fraud Survey 日本企業の不正に関する実態調査*.

<https://assets.kpmg.com/content/dam/kpmg/jp/pdf/2022/jp-fas-fraud-survey.pdf>

梁静, 山田誠二, 寺田和憲, 2015. オンラインショッピングにおける商品推薦エージェントの外見とユーザの購買意欲との関係. *ヒューマンインタフェース学会論文誌* Vol.17 No3, 63-72 307-316.

岡田佳子, 2021. 学生からみたオンライン授業のメリットとデメリット-オンライン環境下のアクティブラーニングに焦点を当てて-. *長崎大学教育開発推進機構紀要*第 11 号.

津村賢宏, 山田誠二, 2021. 擬人化エージェントの自己開示による人の共感の促進. 2021 年度人工知能学会全国大会.

鶴田まなみ, 犬飼圭吾, 2019. ラボ実験を用いた時間制限下の不正行為の検証.

Urs Fischbacher, Franziska Föllmi-Heusi, 2013. Lies in Disguise-An Experimental Study on Cheating. *Journal of the European Economic Association* 11-3, 525-547.