

人の善悪についての評価の形成と変化の非対称性の分析

Analysis of asymmetry in the formation and change of good and evil evaluations on the avatar's personality

中央大学大学院 理工学研究科 ビジネスデータサイエンス専攻

博士課程前期課程 22N710004C 江藤大貴

1. はじめに

現代社会において、差別や偏見に関連する問題が顕著化している。偏った価値判断により行動をする場合、時として適切性に欠く判断や意志決定がなされ、特定のグループや属性に対して過剰に肯定的または否定的な判断をくだす可能性がある。また、自身のバイアスを自覚し改善の意欲を保持していても、是正するよう常に意識した行動は難しい場合がある。これにより、公正で包括的な環境を阻害し、組織や社会の健全な発展を妨げる可能性がある。故に、そのような先入観について理解し対策を講じることが必要である[1][2]。

本研究では特に、人の善悪についての評価に着目する。善悪に対しての先入観による偏見は、人々の意思決定や行動に影響を与える可能性がある。例えば、刑を終えて出所した人に対する偏見は大きい。故に本人に強い更生の意欲がある場合でも、正当な評価がされずに、就職や住居の確保が非常に困難なことがある。また、例えば人種などに基づいた、特定の集団に対して犯罪や違法行為が結び付けられ、刑事司法制度において不公平な判決が生じる可能性がある[3]。

また一般的に、利益よりも損失の心理的効果のほうが大きいという、負効果という現象がある。これにより、道徳的な価値判断において悪い印象が残りやすく良さと悪さの評価は対称でない能力が考えられる。善悪に関係する評価は、損得関係のない単純な確率的な事象のようなものに行う評価とは異なっていると考えられている[4]。このことは上述した元受刑者への偏見のような、不平等な評価につながりかねない。

これらの例が示すように先入観は評価に影響を与えている恐れがある。そこで本研究では、人の善悪についての評価の形成及び変化の様子について検証する。その際に人の善悪といった道徳的判断が入り込まない、単純な確率的な事象と比較する。これにより人の善悪についての評価と、善意や悪意とは無関係の現象への印象の評価との非対称性について実験的に調査する。本研究で主題とする「非対称性」とは、人の善悪についての評価と、善意や悪意とは無関係の現象への印象の評価との間の形成及び変化の様子の違いのことである。

2. 関連研究

石橋[5]は、人が何かを評価する時に目立った特徴に引きずられ、他の特徴について影響を及ぼして評価が歪む現象を、定量的に測定するモデルを提案した。評価者へのアンケート調査を行うことで評価データを収集し、それぞれの評価の歪

む程度の特徴を記述することを可能にした。

北折ら[6]は、殺人事案における裁判員の心証についての研究を行った。検察官や弁護士の発言順を変化させると、裁判員の被告に対する心証が変化するという結果が出ている。

Kojimaら[7]はチーム内での双方向的な共感を促進するため、脳活動を測定することで、道徳意識と自己構造傾向の関係を調査した。実験参加者は道徳に関連する、助けるまたは妨害する行動の場面の映像刺激を視聴した。

これらの関連研究において、善悪の評価の形成と変化に着目した研究は行われていない。また、道徳的判断が入り込まない現象への印象の評価との比較は行われていない。

3. 実験方法

3. 1. 善悪の評価に関する実験

善悪の印象を評価する、30秒のタスク映像を作成した。タスク映像は、乳幼児でも認知できる単純な映像刺激であれば、殆どの人が認知できる映像であると考え、Surianら[8]の研究に用いられた映像刺激を参考にし、Microsoft PowerPointを用いて新たに作成する。良い行動として、ある図形が坂を登ろうとしているところを別の図形が助ける映像(以下「G」)を13種類作成した(図1)。悪い行動として、ある図形が坂を登ろうとしているところを別の図形が妨害する映像(以下「B」)を13種類作成した。13種類の映像はそれぞれ坂の形や、助けられるまたは妨害される図形の色が異なる。

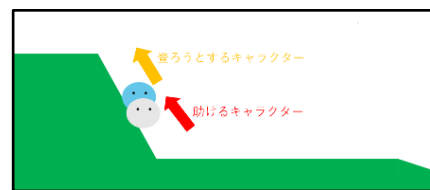


図 1 タスク映像(G)の例の一部

作成した合計26種類のタスク映像を1度ずつ用いて、(1)から(6)の6つのシーケンスを作成した。なお、それぞれの全ての映像では、助けるまたは妨害する図形の色は同一である。

(1)Gを最初に3回提示した後に、Bを次に1回提示し、Gを最後に1回提示するもの《GGGBG》

(2)Bを最初に3回提示した後に、Gを次に1回提示し、Bを最後に1回提示するもの《BBBGB》

(3)Gを5回提示するもの《GGGGG》

(4)Bを5回提示するもの《BBBBB》

(5)Gを最初に1回提示した後に、Bを次に1回提示し、Gを最後

に1回提示するもの《BGB》

(6)Bを最初に1回提示した後に、Gを次に1回提示し、Bを最後に1回提示するもの《BGB》

20代の男性19名、女性9名の合計28名を実験参加者とし、アンケート調査を行った。実験参加者ひとりに対し、(1)から(6)の6つのシーケンスをモニターで示した。同色の図形が同じシーケンス内に登場した場合、前に登場した図形と同じキャラクターだと説明した。これは、助けるまたは妨害する図形がシーケンス内で同一のキャラクターであると実験参加者に認識させ、評価を定着させやすくするためである。また、それぞれのシーケンスにおいて、それぞれのタスク映像を提示し終えた後には7件法のアンケート(表1)を実施した。表1内の亀甲括弧内には「黒色」や「灰色」といった、シーケンスにおいて助けるまたは妨害する行動をとる図形の色が入る。

表 1 善悪の評価に関するアンケート

No.	質問文	回答 (7件法)
問1	映像に良い行動が表れているか?	全く表れていない—完全に表れている
問2	映像に悪い行動が表れているか?	全く表れていない—完全に表れている
問3	()のキャラクターは? (良さ)	全く良いキャラクターではない—完全に良いキャラクターである
問4	()のキャラクターは? (悪さ)	全く悪いキャラクターではない—完全に悪いキャラクターである

3. 2. 確率に関する実験

同じ実験参加者に対し、別の実験を行う。タスク映像はMicrosoft Excelで作成したコイントスを模した映像である。コインを模した楕円が黄色の長方形の中にある。楕円はコインの表(head)を模した「H」、それよりも薄い色のコインの裏(tail)を模した「T」と書いてあるものがそれぞれある。どの映像も、はじめは「H」と書かれた楕円が表示されている。マクロ機能により、同じ位置に配置された「H」、「T」と書かれた楕円が約0.1秒間隔で切り替わり続け、約3秒後にどちらかの楕円が表示される(図2)。同時に楕円外側の黄色の長方形が橙色に変化する。更に「H」と「T」のどちらが表示されたか、シーケンス内で何回目のタスク映像であったかを表す数字と共に出力される。ただしランダムで「H」または「T」が表示されているのではなく、実験参加者には分からない状態でどちらが表示されるのか定められている。3.1節に示した(1)から(6)のシーケンスのGをHに、BをTにそれぞれ置き換えた、(i)から(vi)の6つのシーケンスを作成した。それぞれのシーケンスで、HとTそれぞれで楕円の色は同じである。

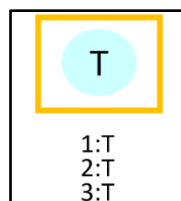


図 2 コイントス映像切り替わり後の例

実験参加者ひとりに対し、(i)から(vi)の6種類のシーケンスをモニターに表示して示した。実験参加者には、タスク映像がコイントスを模していること、「H」がコインの表側で「T」がコインの裏側であること、表と裏が出る確率はそれぞれ1/2ずつである可能性もどちらかが出やすい可能性もあるこ

とを伝えた。それぞれのタスク映像を提示し終えた後には「再び同じコインを投げた際に表が出る確率が0から1の間でいくつだと考えるか」についてアンケートを実施した。コインの表裏が出る確率は1/2ずつでないことを伝えているため、実験参加者は主観的な確率を回答する。主観的な確率とは、ある事象に対して人間が考える信念や傾向の度合いとする。回答する際の参考に、0から100までの5刻みのパーセント表示とその分数表記、分数表記した際に分母が2, 3, 4, 6, 8になる確率及びパーセント表示の表を提示した。実験参加者には、必ずしもその中から選択する必要はないと伝えた。

4. 結果

4. 1. 善悪の評価に関するアンケートの結果

(1)から(6)のシーケンス中の全てのタスク映像後に行った、問1から問4のアンケート結果それぞれについて、帰無仮説を「各データの分布は正規性がある」としたShapiro-Wilkの正規性の検定を行う。その結果、いずれも有意水準5%でアンケート結果に正規性がないと言えた。本実験のデータは対応のあるものであるため、Wilcoxonの符号付き順位検定を有意水準5%で、以下に示す(a)から(h)の8通りの箇所のアンケート結果同士において問1から問4全てについて行う。これにより、(a)から(h)の8通りの比較箇所のアンケート結果同士に差があるか否か、即ち変化があったかどうかを比較する。(a)、(b)では、1度の異なった種類のタスク映像前後で、評価が変化したかを比較する。(c)、(d)では1度の異なった種類のタスク映像を挿入するか否かによって、5度の映像の提示の最後の評価に差があるか否かを比較する。(e)、(f)では、(1)の「GGGBG」及び(2)の「BBBGB」での最初に3度繰り返すタスク映像によって評価が定着されるかを調査するため、(1)、(2)の最後の3度のタスク映像のみで、(a)、(b)と同じ比較を行っている。これに差があるか否かによって、3度繰り返すことが必要か否か推測される。(g)、(h)では(c)、(d)同様に、1度の異なった種類のタスク映像を挿入するか否かによって、3度の映像の提示の最後の評価に差があるか否かを比較する。

- (a) 「GGGBG」での、B前後のG
- (b) 「BBBGB」での、G前後のB
- (c) 「GGGBG」での最後のGと、「GGGGG」での最後のG
- (d) 「BBBGB」での最後のBと、「BBBBB」での最後のB
- (e) 「BGB」での、B前後のG
- (f) 「BGB」での、G前後のB
- (g) 「GGGBG」でのBの直前のGと、「BGB」での最後のG
- (h) 「BBBGB」でのGの直前のBと、「BGB」での最後のB

それらの比較する箇所同士のWilcoxonの符号付き順位検定の統計量Z及びP値を表2に示す。なお、統計量Zの右肩に付したアルファベットは「b」が正の順位に基づくもの、「c」が負の順位に基づくものである。表2に示す結果より、問3の(a)、(c)、問4の(a)、(c)、(e)、(g)はそれぞれ有意水準5%で有意差が認められた。

表 2 善悪の評価に関するアンケート検定結果

No.	比較対象	問1		問2		問3		問4	
		Z	P値	Z	P値	Z	P値	Z	P値
(a)	GGGG	-0.832 ^a	0.405	-1.000 ^a	0.317	-2.045 ^b	0.041	-2.215 ^b	0.027
(b)	BBB	-1.106 ^a	0.916	-2.237 ^c	0.813	-1.078 ^a	0.281	-1.231 ^b	0.218
(c)	GGGG, GGGG	-1.807 ^a	0.071	-1.807 ^b	0.071	-3.211 ^c	0.001	-2.737 ^b	0.006
(d)	BBB, BBBB	-0.528 ^a	0.597	-1.87 ^a	0.851	-0.780 ^a	0.436	-0.941 ^a	0.347
(e)	GG	-0.535 ^a	0.592	-1.76 ^a	0.860	-1.728 ^b	0.084	-2.331 ^c	0.020
(f)	BB	-0.707 ^a	0.480	-0.302 ^a	0.763	-1.365 ^b	0.172	-0.728 ^b	0.467
(g)	GGGG, GBG	-0.028 ^a	0.978	-0.540 ^a	0.589	-1.852 ^b	0.064	-2.110 ^c	0.035
(h)	BBB, BB	-0.577 ^a	0.564	-1.292 ^a	0.197	-1.536 ^b	0.125	-1.830 ^b	0.067

4. 2. 確率に関するアンケートの結果

4.1節のデータと比較するため、Wilcoxonの符号付き順位検定を有意水準5%で、以下に示す(I)から(VIII)の8通りの比較箇所のアンケート結果(主観的な確率)同士に差があるか否か、即ち変化があったかどうかを比較する。

- (I) 「HHHTH」での、T前後のH
- (II) 「TTTHT」での、H前後のT
- (III) 「HHHTH」での最後のHと、「HHHHH」での最後のH
- (IV) 「TTTHT」での最後のTと、「TTTTT」での最後のT
- (V) 「HTH」での、T前後のH
- (VI) 「THT」での、H前後のT
- (VII) 「HHHTH」でのTの直前のHと、「HTH」での最後のH
- (VIII) 「TTTHT」でのHの直前のTと、「THT」での最後のT

それらの比較する箇所同士のWilcoxonの符号付き順位検定の統計量Z及びP値を表3に示す。

表 3 確率に関するアンケート検定結果

No.	比較対象	Z	P値
(I)	HHHTH	-2.018 ^b	0.044
(II)	TTTHT	-3.477 ^b	0.007
(III)	HHHTH, HHHHH	-0.886 ^b	<.001
(IV)	TTTHT, TTTTT	-3.135 ^c	0.001
(V)	HTH	-2.713 ^b	0.376
(VI)	THT	-3.200 ^c	0.116
(VII)	HHHTH, HTH	-1.572 ^b	0.002
(VIII)	TTTHT, THT	-2.914 ^b	0.004

表3に示す結果より、(I)、(II)、(III)、(IV)、(VII)、(VIII)はそれぞれ有意水準5%で有意差が認められた。

5. 考察

5.1. 善悪の評価に関するアンケートによる考察

(a)、(c)、(e)、(g)のアンケート結果の平均値はいずれの箇所も、問3は評価4より大きく、問4は評価4より小さかった。故に、実験参加者は想定された通り、絶対的には「良いキャラクターである」、「悪いキャラクターではない」と評価したと考えられる。(b)、(d)、(f)、(h)では、問3は評価4よりも小さく、問4では評価4よりも大きかった。故に、実験参加者は想定された通り、絶対的には「良いキャラクターではない」、「悪いキャラクターである」と評価したと考えられる。

表2に示す結果より問1, 2, 即ち行動自体の良さや悪さの評価は、(a)から(h)のいずれも有意水準5%を満たさなかった。

故に、比較した箇所の映像自体同士の解釈に差があるとは言えず、実験参加者それぞれが映像自体に異なった解釈をしているとは言えない。

(a)における問3, 即ちキャラクターの良さの印象への回答に対し、有意水準5%で有意差が認められた。故に、3度良い行動を行った後に悪い行動を行い、再び良い行動を行っても、良さの評価は変化すると言える。また、(c)における問3への回答に対しても、有意水準5%で有意差が認められた。「GGGG」と、「GGGBG」の最後のGにおける、良さの評価に差があると言える。これらのことから、良い行動を3度繰り返したことによるキャラクターに対する良さの評価は、1度でも悪い行動をすると変化してしまい、再び良い行動をしても評価されずに変化が表れると言えるということである。(e)及び(g)における問3への回答については、有意水準5%で有意差が認められなかった。故に(e)からは、1度良い行動を行った直後に悪い行動を行い、再び良い行動を行っても、キャラクターに対する良さの評価は変化するとは言えない。また(g)からは「GGG」と「GBG」の最後のGにおける、キャラクターに対する良さの評価に差があるとは言えない。つまり(a)や(c)と同様の展開として、1度の良い行動に対してのキャラクターに対する良さの評価は、悪い行動を1度行った後に良い行動を行っても変化するとは言えないということである。

以上の良い行動における問3への結果から、3度の良い行動を繰り返したことにより形成されたキャラクターに対する良さの評価は、1度の悪い行動により変化しうるのだと考えられる。また1度の良い行動では、1度悪い行動をすることで変化するほどの良さの評価さえ形成されないことが示唆される。

(a)及び(c)における問4, 即ちキャラクターの悪さの印象への回答に対し、有意水準5%で有意差が認められた。故に問3と同様に、良い行動を3度繰り返したことによるキャラクターに対する悪さの評価も、1度でも悪い行動をすると変化してしまい、再び良い行動をしても評価されずに変化が表れると考えられる。(e)及び(g)における問4への回答に対し、有意水準5%で有意差が認められた。故に1度の良い行動でのキャラクターに対する悪さの評価は、悪い行動を1度行った後に良い行動を行うと、変化すると考えられる。以上の良い行動における問4の結果より、3度の良い行動を繰り返したことにより形成されたキャラクターに対する悪さの評価は、1度の悪い行動によって変化しうるのだと考えられる。また、良い行動を3度繰り返さずとも1度の良い行動により、1度悪い行動をすることで変化するほどの弱い悪さの評価が形成されると考えられる。

したがって、良い行動における問3, 4への結果から、良い行動に対する良さと悪さの評価は、1度悪い行動をすることで変化しうるのだと考えられる。また、良い行動に対する良さの評価よりも悪さの評価の方が形成されやすいと考えられる。

(b)及び(d)における問3及び問4への回答について、いずれも有意水準5%で有意差が認められなかった。故に、悪い行動を3度繰り返したことによるキャラクターへの良さ及び悪さの評価は、1度良い行動をし、その後再び悪い行動をしても変化

が表れるとは言えないということである。(f)及び(h)における問3及び問4への回答について、いずれも有意水準5%で有意差が認められなかった。故に、1度の悪い行動でのキャラクターへの良さ及び悪さの評価は、悪い行動を1度行った後に良い行動を行っても、変化するとは言えないということである。

以上の悪い行動における問3及び問4への結果から、3度または1度の悪い行動を繰り返したことにより形成されたキャラクターに対する良さ及び悪さの評価は、1度の良い行動によって変化しないほどの強固なものであると示唆される。または、1度だけでなく3度悪い行動を繰り返したとしても、定まった評価が形成されないとも示唆される。

5. 2. 確率に関するアンケートによる考察

(I), (II), (III), (IV)の回答に、有意水準5%で有意差が認められた。(I), (II)のように主観的な確率は同じ面が3度続けて出た後に、1度異なった面が現れ、再び同じ面が現れることで、異なった面の前後に差が生じたと考えられる。これは、1度異なった面が現れたことにより、連続して現れた面が出る確率を、異なった面が出る前より小さく見積もったためであると考えられる。(III), (IV)のように、同じ面が5回連続して出たものと比較する場合にも差が生じたことから、同様に考えられる。(a)及び(c)の間3, 4の、良い行動に対するキャラクターの良さ悪さの評価の変化も、有意差が認められている点から、この主観的な確率の変化と同様であると考えられる。つまり、道徳的判断の介入しない主観的な確率において、3度同じ面が出た後に1度違う面が表れたことで差が生じたように、3度の良い行動を繰り返したことにより形成されたキャラクターに対する良さの評価が1度の悪い行動によって変化したのだと考えられる。すなわち、良い行動に対する評価は、道徳的判断の介入しない主観的な確率と同等に変化すると解釈される。それに対して、(b)及び(d)の間3, 4の悪い行動に対するキャラクターの良さ悪さの評価の変化は、有意水準5%で有意差が認められていない点から、主観的な確率の変化とは異なると示唆される。つまり、主観的な確率のように差が表れるものではなく、悪い行動に対する良さ悪さの評価が強固なものであると示唆される。すなわち、悪い行動に対する評価は、道徳的判断の介入しない主観的な確率と異なった判断をしていると解釈される。

善悪に関係する評価と、以上のような善悪とは無関係な主観的な確率との比較により、善悪の評価に非対称性がある可能性が示された。

(V), (VI)の回答に、有意水準5%で有意差が認められなかった。このことより、主観的な確率は1度ある面が出た後に、1度違う面が現れ、再び同じ面が現れることにより、違った面の前後に差が生じるとは言えない。これは、最初の2回で2種類の面が出ることにより、3度目にどちらかが現れた際にも大きな確率の違いを実験参加者が感じなかったことが示唆される。これは(VII), (VIII)より、3度ある面が出た場合と、毎回前回とは異なった面が現れた場合で比較した際に、主観的な確率に有意な差

が生じたことから、同様に考えられる。

6. 結論

本研究では、人の善悪についての評価の形成及び変化の様子について検証した。その結果、評価の形成及び変化の様子に違いがあることが分かった。良い行動に対する良さ悪さの評価は、1度悪い行動をすることで変化し、良さより悪さの評価の方が形成されやすいことが分かった。悪い行動に対する良さ悪さの評価は、1度の良い行動では変化しないほどに強固であることが分かった。また、善悪の評価と、道徳判断の介入しない主観的な確率との比較により善悪の評価の間に非対称性があることが分かった。

本研究によって明らかとなった、人の評価の偏りを可視化することが期待される。例えばSNSやメールといった自らの発信の中で、無意識的な評価の偏りに相当する箇所を検出するといったことである。どのような箇所がそういった偏りとして存在しているかを検知するためには、本研究で明らかとなった善悪に関する評価の形成や変化、また非対称性といった特徴についての知が必要であると考えられる。これにより、潜在的に存在している認知バイアスについて課題意識を抱いている個人に対し、それを可視化することが可能になると考えられる。更なる将来的な発展として、公平な社会の実現にも貢献すると期待される。

参考文献

- [1]内閣府男女共同参画局：共同参画2021年10月号, no. 149, pp. 2-5, 2021
- [2]内閣府男女共同参画局：令和4年度 性別による無意識の思い込み（アンコンシャス・バイアス）に関する調査研究, pp. 9-23, 2022
- [3]法務省：令和3年版再犯防止推進白書, 全国官報販売協同組合, pp. 46-57, 2021
- [4]奥田秀宇：対人魅力の非対称性と類似性・非類似性の効果, 社会心理学研究第12巻第2号, pp. 97-103, 1996
- [5]石橋貞人：2次因子分析によるハロー効果の測定, 日本経営工学会論文誌, Vol. 56, No. 2, pp. 121-128, 2005
- [6]北折充隆, 小嶋理江：殺人事案における裁判員裁判の心証に関する研究, 金城学院大学論集人文科学編第15巻, 第1号, pp. 9-18, 2018
- [7]Taiyo Sunny Kojima, Kouki Kamada, Toru Nakata, Takashi Sakamoto, Toshikazu Kato: Brain Activity Difference during Watching Social Behavior Helping Other People. In: Jay Kalra (eds) Emerging Technologies in Healthcare and Medicine. AHFE Open Access, Vol. 116, pp. 386-393, AHFE International, USA. 2023.
- [8]Luca Surian, Mika Ueno, Shoji Itakura, Marek Meristo: Do Infants Attribute Moral Traits? Fourteen-Month-Olds' Expectations of Fairness Are Affected by Agents' Antisocial Actions. Frontiers in psychology, Vol. 9, 1649, 2018.