反実仮想条件文のもっともらしさと感情の関係: Counterfactual Potency と条件付き確率による予測
The relationship between plausibility and emotion in counterfactual conditional: Prediction with Counterfactual Potency and Conditional Probability

渡邊 元樹[†], 高橋 達二[†], 中村 紘子[†‡]
Motoki Watanabe,Tatsuji Takahashi, Hiroko Nakamura
[†] 東京電機大学, [‡] 日本学術振興会
School of Science and Enginneering Tokyo Denki University, Japan Society for the Promotion of Science
22rmd40@ms.dendai.ac.jp

概要
本研究は, 反実仮想条件文「もし p だったら q だっただろう」のもっともらしさの評価について, Petrocelli et al.(2011) の主張する Counterfactual Potency $P(p)$*$P(q|p)$ と Over et al.(2007) が提唱している条件付き確率 $P(q|p)$ の二つのモデルのどちらが予測力が高いかを頻度事例を用いて検討した. また, もっともらしさの評価の際の計算式の自由記述式を用いることで, 計算結果だけではなく計算過程からもモデルの検討を行った [1,2].
キーワード：反実仮想条件文, 条件付き確率, 感情 , Counterfactual Potency

1. 反実仮想
反実仮想条件文とは「もし p だったら q だっただろう」といった, 実際には生じなかった出来事について述べた条件文である.「昨日早く寝ていればテストに遅刻しなかったのに」といった反実仮想は日常的に行われており, 因果関係の認知や, 後悔, 安堵, 怒りといった様々な感情状態にも関わることが知られている. 対話型カウンセリング AI などでは, 発言からその人の認知や感情, 因果関係の捉え方を推測することが必要である. 反実仮想は因果関係や感情の認知に関わるため, 人がどのように反実仮想条件文を解釈し, 推論するかを明らかにすることは, 人の推論のメカニズムを明らかにするだけでなく, より自然な対話型 AI の開発に貢献すると考えられる.

2. 反実仮想条件文のもっともらしさ
反実仮想には「昨日夜空を見ていれば, UFO を見つけられただろう」というもっともらしさの低いものもあれば,「信号無視をしなければ, 怪我はしなかっただろう」というもっともらしさの高いものもある. 人の反実仮想条件文「もし p だったら q だっただろう」のもっともらしさの評価のモデルとして次の 2 つのモデルが提唱されている.

1 つは, Petrocelli et al. (2011) による, 反実仮想のもっともらしさの評価は, p の生じる確率 "If likelihood (IL)" $P(p)$ と, p が生じた場合に q が生じる確率 "Then likelihood (TL)" $P(q|p)$ の同時確率 $P(p)$*$P(q|p)$ であるという Counterfactual Potency である. Petrocelli et al. (2011) は「もし, サムが 2 番のドアを選んでいれば, 賞金を獲得していたかもしれない」という反実仮想条件文について, IL である $P(p)$ の高低 (例：サムは 2 番が好き) と TL $P(q|p)$ の高低を操作したシナリオを提示し, 参加者に反実仮想のもっともらしさや, 登場人物の後悔などの感情を評価させた. その結果, 反実仮想のもっともらしさの評価は, $P(p)$ と P $(q|p)$ の両方が高いときに最も高くなり, Counterfactual Potency によって変化することが示された. また, Counterfactual Potency が後悔, 悲しみといった感情の強さを予測することも示されている [1].

一方, Over et al. (2007) は, 反実仮想条件文のもっともらしさの評価は, p が生じた場合の q の条件付き確率 $P(q|p)$ に従うとしている.

Over et al. (2007) は「もし 2 年前にエイズのワクチンができていたら, アフリカの大きな健康危機は避けられたはずだ」という反実仮想条件文が真だと思う確率と, p & q, p & not q, not p & q, not p & not q に対応する 4 つの事例が真だと思う確率を尋ね, 反実仮想条件文の確率は, 条件付き確率 P $(q|p) = P(p$ & $q)$ / $(P(p$ & $q) + P(p$ & not - $q))$ で予測できることを明らかにした [2].

表 1　IL と TL の高低の各条件における, 前件 p と後件 q の組み合わせの頻度

IL/TL	p & q	p & not q	not p & q	not p & not q	IL : $P(p)$	TL : $P(q \mid q)$	CP
H/H	60	20	10	10	80/100	60/80	60/100
H/L	20	60	10	10	80/100	20/80	20/100
L/H	25	5	10	60	30/100	25/30	25/100
L/L	5	25	10	60	30/100	5/30	5/100

渡邊ら (2023) は人の反実仮想のもっともらしさの評価, およびその評価と因果関係の認知や感情状態と関係について, Counterfactual Potency と条件付き確率のどちらのモデルの当てはまりが良いかを検討した.Petrocelli et al. (2011) , Over et al. (2007) の先行研究ではシナリオを用いていたが, 渡邊ら (2023) では,「もし赤色の玉なら, 当たりだっただろう」といった反実仮想条件文と,「赤色で当たりの玉は 30 個 (p & q) 」,「青色で当たりの玉は 30 個 (not- p & q) 」,「青色でハズレの玉は 10 個 (not p & not- q) 」, といった, 頻度事例を用いて, どの事例の頻度を参照したかから, モデルを検証した. その際, 先行研究である Petrocelli et al. (2011) に倣い p & not- q の事例を含まなかった. 実験の結果, 反実仮想条件文のもっともらしさの評価では, IL と TL の交互作用が見られ, Counterfactual Potency を支持する結果となった [1,2,4].

渡邊ら (2023) の実験は, p & not- q の事例を含まなかったが, 現実世界では「赤色でハズレ」といった p & not q に遭遇することも珍しくない. また, p & not q 事例は反実仮想条件文の反例であり, Counterfactual Potency と条件付き確率のどちらのモデルでも反実仮想のもっともらしさを低下させる事例だといえる. そこで, 本研究では, p & not q 事例を含んだシナリオを用いて反実仮想のもっともらしさの評価モデルを検討した. また, 参加者に反実仮想のもっともらしさの確率を答えてもらうだけでなく,「どのように回答したか数式を書くか, 文章で説明してください」と教示した. 記述式の回答の分析を行うことで, 計算結果だけでなく, どのような計算過程で反実仮想のもっともらしさを評価しているかも検討した [1,4]. 実験では,IL である $P(p)$ 事例の高低と, TL である $P\ (q|p)$ 事例の高低を組み合わせた, HH, HL, LH, LL の 4 つ事例のパターンを作成した.　反実仮想条件文のもっともらしさの確率判断について, Petrocelli et al. (2011) の提唱する Counterfactual Potency では, $P(p)$ と $P\ (q|p)$ の両方が反実仮想条件文のもっともらしさの判断に影響すると予測されている. そのため, 本実験において反実仮想のもっともらしさの確率判断は高い順に HH $>$ LH $\fallingdotseq$ HL $>$ LL という結果が予測され, 因果や感情の評価も順番に同様な順になることが考えられる [1]. 一方, Over et al. (2007) の提唱する条件付き確率では, $P(p)$ の高低によらず, $P\ (q|p)$ に基づいて反実仮想のもっともらしさが判断されると予測されている. そのため, 本実験において反実仮想のもっともらしさの確率判断は高い順に LH $\fallingdotseq$ HH $>$ HL $\fallingdotseq$ LL という結果が予測され, 因果関係や感情の評価も同様な順になることが考えられる [2].

3.　実験

本実験は, 2(IL, $P(p)$: 高・低) * 2(TL, $P(q|p)$: 高・低) の 2 要因参加者間計画で行なった. 反実仮想条件文の前件 p が生じる確率の高低と, p が生じた時に後件 q が生じる条件付き確率の高低を, 対応する事例の頻度によって操作した (表 1).

3.1　参加者と手続き

参加者は, クラウドソーシングサイト (CrowdWorks) を用いて募集した 200 名であった.　このうち, 回答に不備のない 162 名のデータを分析に用いた (平均年齢 43.3 歳, 男性 84 名, 女性 78 名). 実験は, オンラインアンケート調査ツール (Qualtrics) を用いて行い, 参加者はブラウザ上に提示されるシナリオを読み回答した.

3.2　実験材料

実験で用いたシナリオは, 登場人物がクジに外れたという状況についてのものである. 登場人物は赤色と青色の 100 個のボールからなるくじを引き,「青色でハズレのボール」を引き, 賞金がもらえなかった. その後, くじの中身である赤と青の個数や, あたりの個数を示された登場人物は,「もし赤色のボールを引いていれば, 当たりだっただろう」という反実仮想条件文

表 2　IL と TL の高低の各条件における, 確率, 因果, 感情の評定の平均値と標準偏差

IL/TL	N	IL の確率	TL の確率	反実仮想条件文	因果関係	後悔の評価	怒りの評価	落胆の評価
H/H	41	0.739(0.162)	0.687(0.138)	0.615(0.196)	5.46(1.10)	4.37(1.39)	5.65(0.990)	5.17(1.60)
H/L	41	0.709(0.211)	0.292(0.168)	0.250(0.121)	5.10(1.53)	3.89(1.46)	5.73(1.14)	4.37(1.74)
L/H	38	0.316(0.072)	0.632(0.290)	0.528(0.305)	4.61(1.57)	3.50(1.75)	5.26(1.45)	5.39(1.37)
L/L	42	0.291(0.058)	0.157(0.069)	0.176(0.143)	4.07(1.49)	3.40(1.47)	4.90(1.12)	4.36(1.81)

表 3　記述式課題の回答カテゴリー

質問	IL	条件つき確率	連言	Counterfactual Potency	その他	双条件
前件 p が生じたかもしれない確率	132	1	8	0	21	0
p の時に q が生じたかもしれない確率	8	103	24	1	26	0
反実仮想条件文が真である確率	13	61	46	7	34	0

を述べる. くじの中身として, 表 1 のように, p(前件) である赤と青のボールの個数と, q(後件) である当たりとハズレの個数の頻度を操作した 4 パターンのうちのいずれかを参加者に提示し, 参加者はこのシナリオについて, 次の 8 つの質問に回答した.

(1) 前件 p が生じたかもしれない確率
(2) 前件 p が生じたかもしれない確率の計算過程
(3) p の時に q が生じたかもしれない確率
(4) p の時に q が生じたかもしれない確率の計算過程
(5) 反実仮想条件文が真である確率
(6) 反実仮想条件文が真である確率の計算過程
(7) 登場人物がどの程度, 後悔, 怒り, 落胆を感じたかの評価
(8) p が q の原因だと思う程度

評定方法として (1), (3), (5) では 0 〜 100% で評定を行い, (7) , (8) では (1: 全く感じていない) 〜 (7: 非常に感じている) の 7 段階で評定した. (2) , (4) , (6) では自由記述式課題で回答を求めた.

4. 結果

IL, TL の高低の各条件における確率, 因果, 感情の評定の平均値と標準偏差を表 2 に示す.

確率判断, 因果, 感情の評定値それぞれについて IL (高・低) * TL (高・低) の 2 要因参加者間分散分析を行った. その結果, IL の確率判断では, IL の主効果が有意であり $F(1,162) = 357.84$, $p < .001$, IL 高条件で IL 低条件よりも IL の確率が高く評価されていた. TL の確率判断では, TL, IL の主効果が有意であり,TL 高条件で TL 低条件よりも TL の確率が高く評価されていた $F(1,162) = 233.15$, $p < .001$. また, IL 高条件で IL 低条件よりも TL の確率が高く評価されていた $F(1,162) = 11.21$, $p < .001$. 反実仮想条件文の確率判断では, TL, IL の主効果が有意であったが交互作用は見られなかった. TL 高条件で TL 低条件よりも TL の確率が高く評価されていた $F(1,162) = 128.08$, $p < .001$.IL 高条件で IL 低条件よりも TL の確率が高く評価されていた $F(1,162) = 6.37$, $p = .013$. 因果関係の評価では, TL の主効果が有意であり, $F(1,162) = 12.73$, $p < .001$, TL が高いほど p と q の間の因果関係を強く評価していた. 後悔, 怒り, 落胆といった感情の評価では IL の主効果が有意であり (後悔の評価: $F(1,162) = 17.52$, $p < .001$, 怒りの評価: $F(1,162) = 7.92$, $p < .001$, 落胆の評価: $F(1,162) = 10.84$, $p < .001$), IL が高い場合, 後悔・怒り・落胆を強く感じたと評価していた.

記述式の回答は複数名の評価者がコーディングし, IL , Coounterfactual Potency , 条件付き確率 , 双条件 , 連言 , その他に回答を分類した. 記述式の回答の分類結果をを表 3 に示す.

記述式課題における回答カテゴリーの頻度に差があるかについて, それぞれの質問においての回答カテゴリー (IL・条件つき確率・連言・Counterfactual Potency) のカイ二乗検定を行った. その結果,IL の確率では, 有意差がみられ ($x^2(1) = 355.14$, $p < .001$), IL の頻度が他の回答カテゴリーよりも多かった.p の時に q が生じたかもしれない条件付き確率の評価では, 有意差がみられ ($x^2(1) = 194.88$, $p < .001$), 条件付き確率の頻度が他の回答カテゴリーよりも多かった. 反実仮想条件文が真である確率では, 有意差がみられ ($x^2(1) = 62.944$, $p < .001$), 連言と条件付き確率の頻度が他のカテゴリーよりも多かった.

5. 考察

本研究では, 反実仮想条件文のもっともらしさの評価, および, 反実仮想が影響するとされる因果関係の認知や, 怒り, 後悔, 落胆といった感情の予測モデルとして, 条件付き確率, Counterfactual Potency のどちらが当てはまりが良いかを, 事例の生起頻度を操作したシナリオで検討した.

実験の結果, 反実仮想のもっともらしさの評価には, IL と TL の主効果は見られたが, 交互作用が見られなかったことから, 条件付き確率の予測と一致する結果となった. また, 因果関係の評価は TL のみが影響し, 条件付き確率が高い場合に, 因果関係を強く評価するという結果となった. 一方, 感情の評価において, IL の主効果のみが有意であり, 前件 p の生じる確率が高いほど, 怒りや後悔, 落胆を強く感じると評価しており, 渡邊ら (2023) の実験と同じ結果が得られた [4].

反実仮想条件文のもっともらしさの評価は, TL の影響が見られ, 条件付き確率を支持する結果となった. IL の効果が見られた理由として, 計算過程から, 連言として反実仮想のもっともらしさを評価した参加者がいることがあげられる.IL と TL の交互作用が見られなかったことから, 本実験では Petrocelli et al. (2011) の提唱する Counterfactual Potency より Over et al. (2007) の提唱する条件付き確率の方がモデルとして当てはまりがよいといえる [1,2].

因果関係の評価には TL のみが関わり, 渡邊ら (2023) と同じ結果となった. 因果関係の反実仮想モデルである, 構造モデルアプローチ David et al. (2013) では, 「p が q の原因」といえるのは p が介入した場合, q に変化が生じるときだとされている. 本実験の参加者も, p が介入した場合に q がどうなるかという, 条件付き確率をもとに因果関係を判断し, p が生じるか (介入するかどうか) は因果関係の評価に影響しなかった可能性が考えられる [3.4].

また, 感情の評価には, IL のみが関わっており, 渡邊ら (2023) と同じ結果となった. 感情の評価には Counterfactual Potency が影響するという先行研究 Petrocelli et al. (2011) と異なる結果となった. この理由の一つとして感情の評定の際にはポジティブな結果を産むかもしれない p (赤玉を選んでいた) かどうかが重要である可能性が考えられる.

6. 展望

本実験では, 反実仮想条件文のもっともらしさの評価の確率モデルの当てはまりとしては, Over et al. (2007) の提唱する条件付き確率の方が当てはまりがよいという結果になった. ただし, 反実仮想のもっともらしさの評価には, IL の主効果もみられ, また因果関係や感情の評価によって, 人が考慮する事例が異なる可能性が示唆された. 反実仮想と反実仮想が関わる事象の評価において, Petrocelli et al. (2011) の提唱する Counterfactual Potency と Over et al. (2007) の提唱する条件付き確率のモデルどちらが人の思考で使われているのか更なる検討が必要と言える [1,2,4].

文献

[1] Petrocelli, J. V., Percy, E. J., Sherman, S. J., & Tormala, Z. L. (2011) Counterfactual potency. Journal of Personality and Social Psychology, Vol. 100, No. 1, pp. 30–46

[2] Over, D. E., Hadjichristidis, C., Evans, J. S. B., Handley, S. J., and Sloman, S. A.: (2007) The probability of causal conditionals, Cognitive Psychology, Vol. 54, No. 1, pp. 62–97

[3] David A. Lagnado , Tobias Gerstenberg , Ro'i Zultan. (2013) Causal Responsibility and Counterfactuals. Cognitive science. Vol. 37, No. 1, pp. 1036-1073

[4] 渡邊元樹 , 高橋達二 , 中村紘子. (2023) 反実仮想条件文のもっともらしさ ： Counterfactual Potency と条件付き確率の比較 . 人工知能学会全国大会 (第 37 回).