

# 追試とメタ分析による因果帰納推論モデルの適合性評価 Assessing the validity of causal induction models by replications and meta-analysis

林 涼太<sup>†</sup>, 市野 弘人<sup>†</sup>, 樋口 滉規<sup>†</sup>, 高橋 達二<sup>†</sup>

Ryota Hayashi, Hiroto Ichino, Koki Higuchi, Tatsuji Takahashi

<sup>†</sup> 東京電機大学

Tokyo Denki University

tatsujit@mail.dendai.ac.jp

## 概要

本研究では、因果帰納モデルのメタ分析で行われた実験の問題点を指摘し、問題点の大きいと考えられる実験から優先的に追試を実施し、追試データを部分的に組み込み再度メタ分析を実施した。結果として、記述的因果帰納モデルである pARIs が人間の因果的直感に関する内部モデルとして合理的であることを示した。また、心理学研究の再現性や線形モデルを用いた個人差分の観点から実験データやプログラム公開の重要性を説いた。

**キーワード：** 因果推論, 帰納推論, メタ分析, 二重過程理論, 認知モデル

## 1. はじめに

人間を含む動物は因果関係を学習し、推論し、利用することで原因の診断や理由の説明、そして未来の予測を行うことで不確実な環境に適応してきた。人間のような少量データからの素早い学習や判断は、因果関係の推論によって支えられていることが議論されており、人間の因果推論を解明するために、因果帰納モデルの適合による計算論的アプローチが挙げられる。これまでにさまざまな因果帰納モデルが考案され、人間の実験データとの適合とその意味が議論されている。ここで我々は因果推論分野におけるモデル研究のベースとなる因果帰納実験の妥当性、信頼性に着目した。

心理学では、1000 回以上引用されている有名な研究を含め、過去の大部分の実験結果に再現性がないことが明らかとなり、現在改革が進んでいる。[11] このような問題は人間を対象に実験を行う認知科学分野においても懸念される問題であり、妥当な解決策を講じる必要がある。対処方法として追試による実験データの取り直しとメタ分析による研究間を跨いだ統合的な分析が挙げられる。

以上の背景を踏まえ、本研究では、過去に行われた、主要な因果帰納実験の問題点を指摘し、その問題点を

修正する形で追試によるデータの取り直し、メタ分析を行った。

なお、Web 上に再現した実験のソースコード、生データ、分析スクリプトは筆者の GitHub リポジトリにて Web 上に公開する<sup>1</sup>ことで、実験の信頼性が担保されることを期待する。このような取り組みにより、今後の因果帰納推論分野におけるモデル研究の質が高まることが望まれる。

## 2. 因果帰納

### 2.1 因果帰納の二段階理論

本研究は大枠として因果帰納の二段階理論 [4] を採用する。この理論によれば、無数の事象に溢れている現実的な環境における人間の因果推論は、まず第一段階として、現実世界を観察することにより関係の強いと思われる事象のペア（原因候補事象と結果事象）が抽出される。ただし、この第一段階のみでは、相関関係と因果関係は区別できない。しかし、一般的に事象間に因果関係があることは、相関関係もあることを含意するため、第一段階である観察では相関関係のある事象のペアの絞り込みのみ目的とする。

次に第二段階として、第一段階により絞り込んだ事象のペアに対し、原因候補事象が実際に結果事象を引き起こしているかを、反事実を仮定することで、候補となる因果関係を実験的に検討する。この反事実を実験的に検討すること、すなわちコントロール可能性の有無が相関と因果の決定的な違いであり、実験的に介入し分析することで因果関係を認識することが第二段階の目的である。

以上のように因果帰納の二段階理論では、観察と介入を順に配置することで、全ての事象のペアに対して介入する必要がなくなり、本来多大なコストのかかる介入・実験の必要性を限定することができる。このよ

<sup>1</sup><https://github.com/MKryo/LS00> <https://github.com/ichi-269/AS95>

うな観点から限られた認知リソースを持つ人間の因果推論を説明する理論として期待されている。

## 2.2 単純因果帰納

$C$  と  $E$  をそれぞれ単一の事象とおき、この2つの事象の共起情報(表1にある4種類の頻度)から、 $C$  から  $E$  への因果関係の強度を形成するのが、本研究で扱うモデルと実験の枠組みである。以下ではこの枠組みを単純因果帰納と呼ぶ。

表1 単純因果帰納の  $2 \times 2$  分割表

	Effect( $E$ )	No Effect( $\neg E$ )
Causal( $C$ )	$N(C, E)$	$N(C, \neg E)$
No causal( $\neg C$ )	$N(\neg C, E)$	$N(\neg C, \neg E)$

## 2.3 因果帰納モデル

この  $2 \times 2$  分割表での(四分点)相関係数  $\phi$  が基本的な共起情報としての因果帰納モデルである：

$$\phi = \frac{P(C, E)P(\neg C, \neg E) - P(C, \neg E)P(\neg C, E)}{\sqrt{P(C)P(\neg C)P(E)P(\neg E)}} \quad (1)$$

二つ目に紹介する  $\Delta P$  モデルは、因果の確率上昇説に基づいたモデルであり、「 $C$  が  $E$  の確率を上昇させる」とは、 $P(E|C) > P(E|\neg C)$  が成り立つことに他ならないという規則に従う。因果帰納モデルの中では最も規範的であるとされ、(2)式で表される：

$$\Delta P = P(E|C) - P(E|\neg C) = \frac{N(C, E)}{N(C)} - \frac{N(\neg C, E)}{N(\neg C)} \quad (2)$$

このモデルの持つ意味としては、実験対象を2群に分け何かしらの操作を施した群と何も操作を施していない群、この2群間の結果事象の生起確率の差を表しており、科学実験の基礎的な考え方を現している。

具体例としてはとある病気に対する薬の効果を検証するために投薬群( $C$ )とプラセボ群( $\neg C$ )における症状の軽減( $E$ )の確率差を見るものなどが挙げられる。 $\Delta P$  は、合理的(規範的)なモデルでありながら人間の振る舞いを予測・記述しないという特徴を持つ。

記述的モデルを目指して導入されたのが二要因ヒューリスティクス( $DFH$ )である。これは、原因から結果の予測可能性  $P(E|C)$  と、その逆向き、結果から

原因の診断可能性  $P(C|E)$  の幾何平均であり、両者が共に高いときのみ、人間が因果関係を強く感じると予測する。

$$DFH = \sqrt{P(E|C)P(C|E)} = \frac{N(C, E)}{N(C)} - \frac{N(C, E)}{\sqrt{N(C)N(E)}} \quad (3)$$

このモデルの合理性の根拠としては、表1上の  $C$  と  $E$  の相関係数  $\phi$  に対して、 $P(C)$  と  $P(E)$  を0に、あるいは  $N(\neg C, \neg E)$  を無限大に発散させた極限であることが挙げられている。[4]

次に高橋らが提案した pARIs (proportion of assumed-to-be rare instances) というモデルを取り上げる。[9] このモデルは  $DFH$  と似ているがより単純であり、次に示す稀少性仮定の下(現実に近い環境下)でうまく機能することが示されており、人間の因果的直感を表現する。

$$pARIs = P(C \wedge E | C \vee E) = \frac{N(C, E)}{N(C \vee E)} \quad (4)$$

## 2.4 稀少性仮定と因果的直感

$DFH$  の導出で相関係数  $\phi$  に対して行なった  $N(\neg C, \neg E)$  を無限大に発散させた極限をとるという操作は、 $C$  と  $E$  の生起確率が極めて小さいという仮定に基づく、これを稀少性仮定という。これは高次認知の確率論的モデリングで極めて良く用いられるもので、我々が因果関係に関して着目する事象は、その生起が不生起よりも稀である、という一種の事前知識であり、最も緩い形では  $P(C) < .5$ 、最も極端な形では  $P(C) \rightarrow 0$  として表現される。また生起が稀であるということは  $C$  と  $E$  が等確率である ( $P(C) = P(E)$ ) という仮定も同時にみたされる。

ここで、現実的な文脈における人間の因果関係についての、判断、推論を考える。現実世界において多くの場合、判断にはリアルタイム性が求められ、データの収集に対して時間も認知リソースも十分に割くことができない。人間はそういった状況でも因果的直感を適切に働かせ、有効に対処する。このような現実的な環境、すなわち稀少性仮定下における人間の因果的直感の内部モデルを pARIs で説明することに合理性があることが議論されている。[10]

### 3. 認知実験の再現性と信頼性

#### 3.1 心理学研究における再現性危機

近年心理学研究の、再現性問題が大きく注目されている。心理学の主要な雑誌に掲載された97の研究の追試を実施した結果、オリジナルと同様の結果が再現されたのは36%であった[3]。このような状況の中、再現性危機 (Replicability Crisis) や信頼性革命 (Credibility Revolution) と呼ばれる心理学研究の全般的な方法論の見直しが行われている。具体的には追試軽視を批判する動き、古典的研究や主要誌掲載研究を追試するプロジェクト、統合的な分析手法の提案などが行われている。[11]

このような問題意識は人間を対象に実験を行う認知科学分野においても、同様であり、妥当な解決策を講じる必要がある。我々は特に過去に行われた人間を対象とした因果帰納実験に着目した。

#### 3.2 主要な因果帰納実験の問題点と対応

人間の因果関係の学習・判断の仕組みを解明するアプローチとして過去に人間を対象にいくつかの因果帰納実験が行われてきた。ここでは特に Hattori & Oaksford [4] により行われた41種類の因果帰納モデルのメタ分析で用いられた主要な因果帰納実験を取り上げ、いくつかの問題点を指摘するとともに妥当な解決策を講じ、その点を修正する形で追試を実施した。

第一に、実験対象者の偏りが挙げられる。表2に示した因果帰納実験では、ほとんどの研究において世界的に有名な大学の学部生または院生を対象としている。このような西洋文化圏のみを対象とした研究の偏りは、心理学分野において WEIRD サンプルングとして批判されている[5]。この問題への対応として本研究ではクラウドワークスを用いオンライン上で実験参加者を募り、かつ実験実施の時間帯をずらすなどして可能な限り無作為なサンプルングを試みた。ただし、日本のインターネット利用者層も WEIRD に近いため、その点については再考の余地があるが、先行実験と比べると比較的無作為であると考えられる。

第二に、実験参加者の少なさが挙げられる。表2に示した通り各実験の参加者は比較的少ない。特に LS00[6] の実験 1,2,3 の参加者が少ないことがわかる。対応として50~60人の参加者に対してLS00の追試を行い実験データの取り直しを行った。これも上記同様クラウドワークスを用い無作為なサンプルングを行った。

第三に、AS95 [1] ではカバーストーリーの原因事象と結果事象が1対1に対応しない。この点がのちに示すメタ分析の実験選択基準の1. にそぐわないという問題ある。対応として、1対1に修正する形で実験を作り直しデータを集め直した。

第四に、実験の生データを公開していない点が挙げられる。公開されているデータは回答の平均値などの統計量のみであり、個人の回答データは公開されていないことが一般的である。そのため既存の結果からでは、線形モデルを使用した個人差分析などの詳細な分析ができない。対応として、本研究で収集した実験データ、分析スクリプト、実験プログラムをweb上に公開した。これにより実験データの信頼性の向上とともに、認知実験で用いたすべてのデータを公開することが、分野全体への貢献につながるという意味で望ましいという考えが定着することを期待する。

表2 Hattori & Oaksford (2007) のメタ分析に含められた因果帰納実験の参加者数と母集団

実験	人数	実験対象者
AS95 [1]	40	カーネギーメロン大学
BCC03.1 [2]	52	カリフォルニア大学
BCC03.3 [2]	31	シェフィールド大学
HO07.1 [4]	39	立命館大学
HO07.2 [4]	50	立命館大学
LS00.1 [6]	28	University College London
LS00.2 [6]	16	フィリップ大学マールブルク
LS00.3 [6]	26	University College London
W03.2 [8]	40	カーディフ大学
W03.6 [8]	43	カーディフ大学

### 4. メタ分析

[Takahashi et al 2019] [7] は過去の実験データに対する包括的なメタ分析を行い、pARIs と DFH の記述的性能比較が行われ、pARIs が DFH に勝るとも劣らないデータの説明力が示された。本研究ではメタ分析に含められた実験の中から、上記で指摘した LS00, AS95 の問題点を修正する形で追試を実施し、取り直したデータに置き換え、再度同様のメタ分析を行った。

#### 4.1 因果帰納モデルのメタ分析における実験の選択基準

Hattori & Oaksford (2007)[4] や Takahashi et al (2019)[7] で行われた因果帰納モデルのメタ分析で採用された因果帰納実験の選択基準を以下に示す。

1. それぞれただ一つの原因候補と注目する結果のみが提示される
2. 主観的な関係の強さが（ほぼ）連続的な値として判断される
3. 実験家が決定した頻度に従って、原因候補と結果が逐次的に提示される
4. タスクは、原因の生起が結果の生起を抑制する抑制原因でなく生起が生起を促進する生成原因を扱うものである

## 5. 実験 1

### 5.1 実験概要

本実験は LS00 の忠実再現と変更を加えた再現の 2 種類の追試を行った。

### 5.2 実験手順

本実験は、作成した Web ページを用いて実施した。Web ページの作成に関しては、フレームワークとして Django を使用、サーバーは Amazon Web Services を用いた。また実験の実施については、クラウドソーシングサイトであるクラウドワークスを用いることで、無作為かつ効率的に実験参加者を募集した。

この実験は単純因果帰納の枠組みのもと実施した。実験参加者は結果事象 E とその原因の候補と考えられる事象 C があるときに C と E が共に起こった事象、C が起こり E が起こらなかった事象、C が起こらず E が起こった事象、C と E がどちらも起こらなかった事象、という 4 種類の共起事象をそれぞれ何度か、決まった回数・ランダムな順序で提示され、そこから C が E の原因であると感じる程度を評価する。4 種類の共起事象をそれぞれ (a,b,c,d) とおいたとき、どの事象が何回提示されるかという各セルの試行数を表 3 に示した。

以下に実験参加者に提示した文章を示す。<sup>2</sup>

### 5.3 提示したカバーストーリー

- これから「化学物質の投与」と「動物の遺伝子の突然変異」という二つの事柄の関係性を調査している研究者を演じていただきます。
- 上記の調査のため、3 種類の動物実験が行われました。以下ではその実験記録を見ていくことにします。

- 3 つの実験それぞれで、60 個体の動物が用いられました。その中で、化学物質を投与された個体と投与されなかった個体がありました。注意すべきこととして、動物では、化学物質の投与の有無に関わらず、自然に遺伝子の突然変異が起こることがあります。60 個体について、しばらくしてから遺伝子を調べ、突然変異が起きているかどうかを確認しました。
- 上記で説明した動物の実験記録を一つ目のマウスの事例から順に、1 匹ずつ提示するので、動物の遺伝子が突然変異するのかもしれないかを予測してください。
- 動物の遺伝子が「変異する」「変異しない」どちらかのボタンを押すと、実際にその動物の遺伝子が増えたのか、変異しなかったのか、答えが表示されます。これをもとに「動物に化学物質を投与することが、動物の遺伝子の突然変異を引き起こしているかどうか」について因果的な関係性の強さを判断していただくということがこの実験の主旨になります。
- 最初は判断に迷うかもしれませんが、回答に迷った時は深く考えず直感的に回答していただいて問題ありません。回数を重ねるごとに迷わず判断できるようになるはずです。

### 5.4 実験 1-1

上記カバーストーリーの通り、ある架空の化学物質をマウス、ウサギ、ハト 3 種類の動物に投与し、その化学物質が動物の遺伝子の突然変異を引き起こすかどうかを観察してもらった。先に動物へ化学物質を投与したかどうかの情報を提示し、その後動物の遺伝子が突然変異したかどうか、すなわち原因事象から結果事象を毎提示予測してもらった。(図 1) 予測の直後に実際にその動物の遺伝子が増えたかどうか予測の答えを提示した。このように、参加者は動物に化学物質を投与した結果を 1 つずつ順に観察し、10 回刺激を観察するごとに化学物質が動物の遺伝子変異に対する因果関係の強さの推定値を計 6 回回答してもらった。化学物質と遺伝子の変異の因果関係の評価は 0 から 100 までの値をスライダーで回答させる形式をとった。

### 5.5 実験 1-2

基本的には実験 1 と同様の形式を取ったが異なる点が 3 つある。一つ目はカバーストーリーにモルモット

<sup>2</sup>LS00 のカバーストーリーを理解し易い形に翻訳している

を追加し計4種類の動物を用い、それに合わせて刺激を4つに増やした。二つ目として、各セルの試行数の合計を56へと変更した。最後に、推定値を聞く回数を6回から3回に減らした。

## 5.6 実験1-3

基本的には実験2と同様の形式を取ったが、唯一の変更点として各セルの試行数の合計を60へ戻した。

以上1-1から1-3まではオリジナルの実験を忠実に再現したものである。次にオリジナルの特異点、すなわち原因事象の表示と結果事象の表示の間に予測を挟む点を変更し、原因事象と結果事象を同時に表示するように変更を加えた追試実験を行なった。

## 5.7 実験1-4

実験1の原因から結果の予測を毎提示行ってもらおうという実験形式を、原因と結果を同時に提示する形式(図2)へと変更した。それ以外は実験1と全く同様の形式とした。

## 5.8 実験1-5

実験2の原因から結果の予測を毎提示行ってもらおうという実験形式を、原因と結果を同時に提示する形式へと変更した。それ以外は実験3と全く同様の形式とした。

## 5.9 実験1-6

実験3の原因から結果の予測を毎提示行ってもらおうという実験形式を、原因と結果を同時に提示する形式へと変更した。それ以外は実験3と全く同様の形式とした。

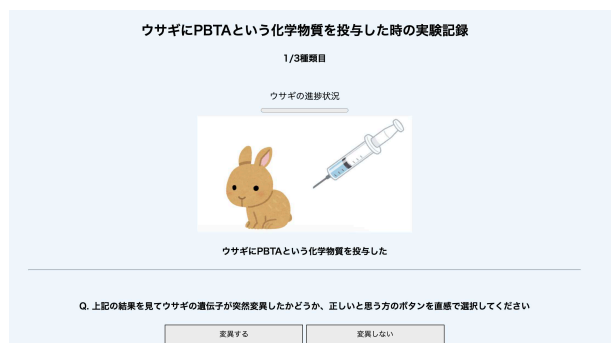


図1 実験画面のサンプル

表3 実験1における刺激のデザイン

刺激	a	b	c	d	N
1	21	9	0	30	60
2	24	6	10	20	60
3	27	3	20	10	60
4	28	0	0	28	56
5	28	0	7	21	56
6	28	0	14	14	56
7	28	0	21	7	56
8	30	0	18	12	60
9	24	6	12	18	60
10	12	18	0	30	60
11	27	3	25	5	60

## 6. 実験2

### 6.1 実験概要

実験2ではAS95の忠実再現と変更を加えた再現の2種類の追試を行い、データを取り直すことを目的として実施した。

### 6.2 実験手順

実験1同様、作成したWebページを用いて実施した。ページ作成に関して使った技術は実験1同様であり、この実験も単純因果帰納の枠組みのもと実施した。オリジナルのAS95は実験参加者1人あたりに80刺激を提示し、実験にかかる時間は約1時間であった。この形式は、ワーキングメモリや認知リソースの限られた人間に対する認知実験としては適切であるとはいえない。また、カバーストーリーとして提示する因果構造が、1対1対応していないという点も単純因果帰納の枠組みから逸脱しているという意味では、適切な形式とはいえない。



図2 実験画面のサンプル

まず、2つの実験ともに一人当たりに提示する刺激数のみを減らし実験時間の短縮を行った。忠実再現として1対1対応しないカバーストーリーをそのまま提示、変更再現としてカバーストーリーの因果構造を1対1対応するように変更した。各刺激を表4に示した。以下に実験参加者に提示した文章を示す。<sup>3</sup>

### 6.3 提示したカバーストーリー

- これから「ラフマ症候群」と、それに対する治療薬に関して調査をしている研究者を演じていただきます。
- ある製薬会社がラフマ症候群の治療薬を13種類開発し、調査のためにそれぞれの薬を全国13箇所の病院に配布しました。
- 治療薬には副作用として発疹が懸念されます。
- 各病院にて病気の患者に対して治療薬の投与が行われ、患者が副作用を発症したかどうか調査されました。
- あなたには、それぞれの薬がどの程度の割合で副作用を引き起こすかを推定してもらいます。
- 副作用の発疹はラフマ症候群の症状とは無関係であること、各病院で投与される薬はそれぞれ別の種類の薬であることに注意してください。

表4 実験2における刺激のデザイン

刺激	a	b	c	d	N
1	3	3	3	3	12
2	9	3	3	3	18
3	3	5	5	5	18
4	9	5	5	5	24
5	3	9	3	3	18
6	5	3	5	5	18
7	5	9	5	5	24
8	3	3	9	3	18
9	5	5	3	5	18
10	5	5	9	5	24
11	3	3	3	9	18
12	5	5	5	3	18
13	5	5	5	9	24

### 6.4 実験2-1

オリジナルのカバーストーリーは以下のように提示されており、結果事象である発疹が、薬の副作用によ

<sup>3</sup>AS95のカバーストーリーを理解し易い形に翻訳している

るものか、ラフマ症候群の症状によるものか一意に定まらないため実験参加者の混乱を招く。提示した実験画面の例を図3に示す。

- 「治療薬には副作用として発疹が懸念されます。ただしラフマ症候群自体の症状としても、発疹がよくあらわれると言われています。」
- ここではカバーストーリーを以下のように薬の副作用によってのみ発疹が起こることを示し、因果構造を1対1に修正し提示した。
- 「治療薬には副作用として発疹が懸念されます」



図3 実験2画面のサンプル

### 6.5 実験2-2

オリジナル同様のカバーストーリーを提示した

## 7. 結果

モデルの評価値と実験参加者の回答の平均値との決定係数を表5,6に示す

表5 LS00 (実験1)

	オリジナル	忠実追試	変更追試
$\Delta P$	.77**	.40*	.58**
$DFH$	.80**	.84**	.89**
$pARIs$	.80**	.80**	.87**

Note: \*\* $p < .01$ , \* $p < .05$ . p値は無相関検定によるもの

表6 AS95 (実験2)

	オリジナル	忠実追試	変更追試
$\Delta P$	.78**	.54**	.57**
$DFH$	.91**	.93**	.94**
$pARIs$	.89**	.92**	.94**

Note: \*\* $p < .01$ , \* $p < .05$ . p値は無相関検定によるもの

## 7.1 線形混合効果モデルを用いた個人差分分析

線形混合効果モデル (Linear Mixed-Effects Model) を用いて、実験参加者個人をランダム効果とし、個人差を考慮した分析を行った。従属変数を人間の回答、説明変数を各モデルの評価値として、実験参加者個人ごとに回帰させた。(人数分の回帰直線を考える)

個人間の回帰直線の傾きと切片の標準偏差を表7に示す。なお、LS00に関しては、実験を小分けにしており3~4刺激ごとに参加者が変わってしまうため、個人をランダム効果とした線形混合効果モデルを適用することはできない。従ってここではAS95の追試データのみを対象とする。

pARIsの評価値を説明変数としたときおよび、DFHの評価値を説明変数にしたときこのモデル間では、回帰直線のばらつきの個人差にあまり違いはないが、 $\Delta P$ の評価値を説明変数にした時、回帰直線の個人間の標準偏差が小さい傾向にあることが読み取れる。これは、人間の回答に対する説明変数として、pARIsやDFHを採用すると個人差が大きく、 $\Delta P$ を採用すると個人差が小さくなる傾向にあることを示す。この結果はpARIs、DFHの直感的、創造的傾向、 $\Delta P$ の熟慮的、論理的傾向が現れているのではないだろうか。

しかし、表8に示すように、個人ごとに切片と傾きがそれぞれ独立に決まるモデルを用い回帰すると標準偏差の差はあまり読み取れない。人間の回答と因果帰納モデルの評価値との回帰直線において、回帰係数と切片の個人間のばらつきがどんな意味を持つのか、また線形モデルの違いによる影響が明確にわかっているわけではないため、今後より深く慎重に考えていく必要があると思われる。そのためにも、誰がどの環境でも個人差を考慮した分析ができるように、個人レベルの実験データを公開することが必要である。

表7 Random effects (as95)

実験 2-1		実験 2-2			
	$\beta_0$ -SD	$\beta_1$ -SD			
$\Delta P$	11.46	16.94	$\Delta P$	10.14	13.82
DFH	27.46	39.76	DFH	18.86	30.42
pARIs	21.96	40.69	pARIs	15.24	30.96

Note: 実験参加者間の回帰直線において、 $\beta_0$ -SD は切片の標準偏差、 $\beta_1$ -SD は傾きの標準偏差を表す。

表8 Random effects (as95) 傾きと切片が独立して決まる場合

実験 2-1		実験 2-2			
	$\beta_0$ -SD	$\beta_1$ -SD			
$\Delta P$	11.46	16.94	$\Delta P$	10.14	13.82
DFH	11.59	.00	DFH	9.75	9.65
pARIs	11.70	8.81	pARIs	10.01	14.29

Note: 実験参加者間の回帰直線において、 $\beta_0$ -SD は切片の標準偏差、 $\beta_1$ -SD は傾きの標準偏差を表す。切片と傾きがそれぞれ独立に決まるモデルを使用

## 7.2 実験1と2の統合

実験1と2の結果を過去のメタ分析で用いられた実験と統合し各モデルについて固定効果モデルを用いて相関係数を、点推定、区間推定した。結果を表9に示す。追試データを部分的に差し替えることによりpARIsの相関係数が.90から.94へ、DFHの相関係数は.91から.92へ向上した。一方 $\Delta P$ の相関係数は.72から.54へ減少した。これは、モデルの性質上望ましい結果であると考えられる。

## 8. 考察

一般的に、ワーキングメモリ容量や認知リソースの限られた人間が扱うデータはかなり小さい。過去のシミュレーション研究からpARIsは少数のサンプルしか与えられない場合においても、早く正確に計算可能であることがわかっている。このような人間の認知的特徴を考慮した現実的な環境において有効に働くという理由から人間の因果的直感に関する内部モデルとしてのpARIsの合理性が示唆されてきた。[10]

本研究では、メタ分析に含まれる因果帰納実験の問題点を指摘し、問題点の大きい実験から優先的に追試を実施した。追試により取り直したデータを部分的に入れ替え、メタ分析を実施した結果pARIsのフィットが向上した。

この結果から、認知実験の側面から人間の因果的直感に関する内部モデルとしてpARIsの合理性を示すことができた。しかし、pARIsのフィットが良くなるように恣意的に実験データを差し替えたようにも捉えられる。本研究ではそのような誤解を避けつつ実験の信頼性を担保するためにも、作成した実験のプログラムと収集したデータ、分析スクリプトを公開し、誰がどの環境でも実験の正当性を確認できるようにフォーマットを整えて公開した。このように認知実験に用いたすべてのプログラムとデータを公開しているケースはまだ少ないように思われる。認知科学分野

表9 メタ分析

	AS95	BCC03.1	BCC03.3	HO07.1	HO07.2	LS00	W03.2	W03.6	$r$	$L$	$U$
オリジナル											
pARIs	.89	.96	.94	.93	.99	.80	.78	.88	.90	.86	.93
DFH	.91	.95	.91	.93	.96	.80	.69	.80	.91	.87	.93
$\Delta P$	.78	.84	.70	.50	0	.77	0	NA	.72	.63	.80
忠実追試											
pARIs	.92	-	-	-	-	.80	-	-	.93	.88	.96
DFH	.93	-	-	-	-	.84	-	-	.91	.85	.95
$\Delta P$	.54	-	-	-	-	.40	-	-	.50	.25	.70
変更追試											
pARIs	.94	-	-	-	-	.87	-	-	.94	.89	.96
DFH	.94	-	-	-	-	.89	-	-	.92	.87	.96
$\Delta P$	.57	-	-	-	-	.58	-	-	.54	.32	.71

Note: - はオリジナルのメタ分析と同様の値であることを表す。  $r$  は相関係数の点推定,  $L, U$  は相関係数の区間推定値を表す。

全体において、今後さらなる追試研究と情報の公開による信頼性の確保が求められる。

## 9. 結論

本研究では、過去の因果帰納実験の問題点を指摘し、問題点の多い実験から優先的に追試、統合的に分析することで、認知実験の側面からも pARIs の合理性を示すことができた。

pARIs, DFH は共に人間の内部モデルの説明として優れていることが明らかになりつつある、今後はベイジアンネットワークなどの因果探索アルゴリズムへの組み込み、そのために  $N \times N$  または連続的な枠組みへの理論の拡張が検討されるべきである。

また、認知実験分野においてソースコードを公開し、誰がどの環境でも実験の確からしさを確認できるようにすることの重要性を、線形モデルを用いた個人差分析、再現性や信頼性の観点から示すことができたのではないかと考える。このような取り組みにより、因果推論分野においてモデル研究の質の向上が望まれる。

## 文献

- [1] John R Anderson and Ching-Fan Sheu. Causal inferences as perceptual judgments. *Memory & cognition*, 23(4):510–524, 1995.
- [2] Marc J Buehner, Patricia W Cheng, and Deborah Clifford. From covariation to causation: a test of the assumption of causal power. *Journal of experimental psychology: learning, memory, and cognition*, 29(6):1119, 2003.
- [3] Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716, 2015.
- [4] Masasi Hattori and Mike Oaksford. Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. *Cognitive science*, 31(5):765–814, 2007.
- [5] Joseph Henrich, Steven J Heine, and Ara Norenzayan. The weirdest people in the world? *Behavioral and brain sciences*, 33(2-3):61–83, 2010.
- [6] Klaus Lober and David R Shanks. Is causal induction based on causal power? critique of cheng (1997). *Psychological review*, 107(1):195, 2000.
- [7] Tatsuji Takahashi, Kuratomo Oyo, Akihiro Tamatsukuri, and Kohki Higuchi. Correlation detection with and without the theories of conditionals: A model update of hattori and oaksford (2007). In *Logic and Uncertainty in the Human Mind*, pages 207–226. Routledge, 2020.
- [8] Peter A White. Causal judgement as evaluation of evidence: The use of confirmatory and disconfirmatory information. *The Quarterly Journal of Experimental Psychology*, 56(03):491–513, 2003.
- [9] 高橋達二 and 菅野宏明. 因果帰納の対称性と非対称性. *日本認知科学会第 27 回大会発表論文集*, pages 199–200, 2010.
- [10] 高橋達二 and 大用庫智. 対称性推論のモデルとしての「双条件付確率」と小数サンプルからの因果帰納推論. *JCSS Japanese Cognitive Science Society*, pages 141–148, 2014.
- [11] 池田功毅 and 平石界. 心理学における再現可能性危機: 問題の構造と解決策. *心理学評論*, 59(1):3–14, 2016.