

因果推論におけるパワー PC 理論について

斎藤 元幸・嶋崎 恒雄

1. 序 論

原因と結果の関係である因果関係を直接観察することはできないが、それによって様々な事象が結び付けられている。観察可能である事象の生起から推論することにより、観察不可能である因果関係は明らかにされる。ヒトは因果の知識を利用することで様々な事象の説明・制御・予測を行っている。風邪を引いた時には、なぜ風邪を引いたか考える。そして、体調をコントロールしようと薬を飲み、風邪が治ることを期待する。このような認知機能が可能となるためには、因果関係の把握が必要不可欠である。

ヒトがどのようにして因果関係を知り得るかという因果の認識の問題は哲学だけでなく、心理学の様々な分野でも数多くの議論がなされてきた。知覚心理学では因果知覚 (causal perception)、認知心理学では因果帰納 (causal induction) や因果推論 (causal inference)、学習心理学では随伴性判断 (contingency judgment) や随伴性学習 (contingency learning)、社会心理学では原因帰属 (causal attribution) といったテーマで研究が進められてきた。これらは研究領域やその目的に基づいて分類されているが、いずれの実験事態においても、実験参加者に対して何らかの情報を与え、因果関係の判断を行わせる手続きが用いられる (嶋崎, 1995)。本稿では、因果推論と呼ばれる分野に焦点を当て、その一端を紹介することを目的とする。

因果推論の実験事態では、事象の生起情報は何らかのカバーストーリーを通して与えられる。例えば、病気に対する薬の効果を検証するというカバーストーリーでは、薬と病気という 2 つの事象を観察することが求められる。薬を

飲んで病気が治った事例や、薬を飲まなかったが病気が治った事例を観察して因果関係を判断させる。因果関係の判断もカバーストーリーに応じた文脈でなされ、上記の例では病気に対する薬の効果を評定することが求められる。

因果推論のメカニズムを解明しようと、様々なアプローチから研究が盛んに行われてきた (see Penn & Povinelli, 2007 ; Sawa, 2009 ; Shanks, 2007 for reviews)。因果推論研究の初期の頃から大きな影響力を持っているのは、動物の条件づけ研究に端を発する連合的アプローチ (associative approach) である。因果推論の実験事態で 2 つの事象を観察させることは、古典的条件づけの条件刺激 (conditioned stimulus : CS) と無条件刺激 (unconditioned stimulus : US) の対呈示と構造的に類似していることから、条件づけの理論を用いて因果推論の過程を説明する試みがなされてきた。因果推論のモデルに適用されている中で代表的なモデルは Rescorla-Wagner モデル (Rescorla & Wagner, 1972) である⁽¹⁾。このモデルによると、因果関係の学習は原因と結果の連合強度の変化であり、その変化は次のように表現される。

$$\Delta V_i = \alpha_i \beta (\lambda - \sum V_i) \quad (\text{式 1})$$

ΔV_i は原因と結果の連合強度の変化量、 α_i は原因についての学習率パラメーター、 β は結果についての学習率パラメーターをそれぞれ表す。原因と結果が共生起した時は $\lambda = 1$ となり、連合強度は増加する。逆に、原因のみが生起した時は $\lambda = 0$ となり、連合強度は減少する。 $\sum V_i$ は既に形成されている連合強度を表し、学習の初期などでこの値が小さいほど連合強度の変化量は大きくなる。

上記の連合的アプローチと対立するのが法則基礎的アプローチ (rule-based approach) である。法則基礎的アプローチでは、事象の生起情報から何らかの計算法則で因果関係を引き出す過程として因果推論が捉えられている。事象の生起情報は Figure 1 に示した随伴性テーブルで表現される。4 つのセルは

(1) 因果推論のモデルとして適用を試みる条件づけの他の理論として、Pearce モデル (Pearce, 1987) やコンパレータ仮説及び拡張版コンパレータ仮説 (Denniston, Savastano, & Miller, 2001 ; Miller & Matzel, 1988) が挙げられる。

		結果事象 E	
		e	$\neg e$
原因事象 C	c	a	b
	$\neg c$	c	d

Figure 1 随伴性テーブル

各々の事象の共生起の頻度を表している。例えば、セル a は原因と結果が共生起した事例を、セル d は原因と結果が生起しなかった事例を表している。この随伴性テーブルで表現される事象の生起情報に対して何らかの計算が行われる。計算法則には、ヒューリスティックや 2 変量の関係を表す統計量など様々なものが考案されている (cf. Hattori & Oaksford, 2007)。法則基礎的アプローチに

おける代表的なモデルは ΔP モデル (Jenkins & Ward, 1965) であり、次のように定義される。

$$\Delta P = P(e|c) - P(e|\neg c) = \frac{a}{a+b} - \frac{c}{c+d} \quad (\text{式 2})$$

$P(e|c)$ は原因が生起した時に結果が生起する確率を、 $P(e|\neg c)$ は原因が生起していない時に結果が生起する確率を表している。 ΔP はこれら 2 つの条件付き確率の差分をとったものである。 ΔP は -1 から 1 までの値をとり、値によって意味合いが異なる。 $\Delta P > 0$ となると原因は結果を発生させ、 $\Delta P < 0$ となると原因は結果を抑制し、 $\Delta P = 0$ となると原因と結果は無関係となる。

連合的アプローチの Rescorla-Wagner モデルと法則基礎的アプローチの ΔP モデルの間には密接な関係がある。結果についての学習率パラメーター β が一定の事態⁽²⁾において、Rescorla-Wagner モデルの漸近値と ΔP の値は一致することが知られている (Chapman & Robbins, 1990 ; Wasserman, Elek, Chatlosh, & Baker, 1993 ; see Danks, 2003 for detailed analysis)。

これら 2 つのアプローチは、事象の生起情報に重点をおいていることから共変動アプローチ (covariation approach) と言われる。共変動アプローチに

(2) 一般的に β は一定の値ではなく、US 呈示時の $\uparrow \beta$ は US 非呈示時の $\downarrow \beta$ よりも大きく設定される。これにより、学習よりも消去に時間が掛かることが説明される。

における問題点は、共変動と因果が区別されていないことである。つまり、共変動がいつも因果関係を示しているとは限らない。例えば、気圧計の値と天気の状態は共変動にあるが、気圧計を操作しても雨が降ることはない。低気圧という共通の原因が背後に存在するために生じる擬似相関であるため、このような操作は効果をもたない。

共変動と因果の区別はパワーアプローチ (**power approach**) によって説明される。パワーアプローチでは、共変動と因果を区別する機能を持つ何らかの知識が仮定されている。ここで言う知識とは、原因が結果を発生させるパワー、または原因と結果の間に存在するメカニズムについての知識であり、これらの知識が機能しない限り因果関係が認知されることはない。**Cheng (1997)** はこれらの知識を総称して、因果効力 (**causal power**) と呼んでいる。しかし、このアプローチには 2 つの問題点が存在する。パワーアプローチは共変動アプローチと異なって計算不可能であるため、結果を十分に予測することができない。また、パワーアプローチによる説明は循環論に陥りやすい。パワーアプローチは、因果推論をパワーやメカニズムなどの知識によって説明するが、これらの知識がどのように獲得されるかについて何も述べていない。知識の獲得を問題とする時、再び因果推論がどのようになされるかという問題を考えなければならない。

共変動アプローチとパワーアプローチ⁽³⁾は互いに問題点を抱えており、その解決には両者の統合が必要不可欠である。そこで本稿では、これら 2 つのアプローチの統合を試みるパワー PC 理論を概説し、その妥当性を検討した研究を展望する。

(3) **Sloman (personal communication, August 2009)** は共変動アプローチを差異産出理論 (**make a difference theories**)、パワーアプローチを生成的理論 (**generative theories**) と呼んでいる。

2. パワー PC 理論と結合パワー PC 理論

2-1. パワー PC 理論

Cheng (1997) は共変動アプローチとパワーアプローチにおける問題点を解決するために、両者を統合するパワー PC 理論⁽⁴⁾ (power PC theory) を提唱した。因果効力を定式化することによって、共変動と因果の違いを明確にし、モデルを計算可能にした。このモデルによると、ヒトは結果を発生させる力を持つ原因と結果を抑制する力を持つ原因だけが結果に影響を及ぼすと信じており、これらの因果効力を用いて観察した事象を説明しようとする。結果を発生させる力は発生的因果効力 (generative causal power) と呼ばれ、結果を抑制させる力は抑制的因果効力 (preventive causal power) と言われる (Cheng, 1997)。

因果推論を因果効力などの理論的存在を用いて説明する、パワー PC 理論 (Cheng, 1997) や因果モデル理論 (causal model theory: Waldmann & Holyoak, 1992; Waldmann, 1996; Waldmann & Martignon, 1998; Waldmann & Hagmayer, 2001; Waldmann, Hagmayer, & Blaisdell, 2006) などの試みは因果的アプローチ (causal approach) と呼ばれ、連合的アプローチや法則基礎的アプローチとは区別される。

結果 e に対する原因 i の発生的因果効力は q_i で表され、抑制的因果効力は p_i と表現される⁽⁵⁾。これらは原因 i が存在するときに結果 e を発生または抑制する確率である (i.e., $0 \leq q_i \leq 1$, $0 \leq p_i \leq 1$)。因果効力は理論的存在であるため、その値は間接的に推定される。他の原因が全く存在しないとき $P(e|i) = q_i$ となるが、一般的には一致することはなく、他の原因 a (ここでは i 以外

(4) PC は確率対比 (probabilistic contrast) の略称である。

(5) Cheng (1997) ではどちらの因果効力も p で表現されているが、Novick & Cheng (2004) や Lu, Yuille, Lijeholm, Cheng, & Holyoak (2008) は発生的因果効力を q 、抑制的因果効力を p と表しており、本稿ではそれらに準ずる。

の原因) について考慮する必要がある。因果効力を推定する際に、考慮の対象となる事例は焦点セット (focal set) と呼ばれる。

2-2. 発生的因果効力

以下の 4 点は、発生的因果効力の算出における仮定である。

- (1) 原因 i と原因 a は独立に結果 e に影響する。
- (2) 原因 a によって結果 e が発生することはあるが、抑制されることはない。
- (3) 原因 i と原因 a の因果効力はそれらの生起頻度とは無関係である。
- (4) 原因が無いときに結果 e が生じることはない。

$\Delta P_i \geq 0$ となる時、上記の仮定に基づいて発生的因果効力 q_i が算出される。算出式を理解しやすいように、まず始めに因果効力で条件付き確率や ΔP_i の説明を行う。その後、発生的因果効力について説明する。結果 e が生起する確率 $P(e)$ は、原因 i が q_i で結果を発生させる確率と原因 a が q_a で結果を発生させる確率を加算し、両者の共通部分を引くことにより算出される。

$$P(e) = P(i) \cdot q_i + P(a) \cdot q_a - P(i) \cdot q_i \cdot P(a) \cdot q_a \quad (\text{式 3})$$

原因 i が存在するという情報を付加する⁽⁶⁾ことによって、条件付き確率 $P(e|i)$ が求まる。

$$P(e|i) = q_i + P(a|i) \cdot q_a - q_i \cdot P(a|i) \cdot q_a \quad (\text{式 4})$$

同様に、 i 原因が存在しないという情報が付加されることで、条件付き確率 $P(e|\neg i)$ が求まる。

$$P(e|\neg i) = P(a|\neg i) \cdot q_a \quad (\text{式 5})$$

式 4 と式 5 の差分をとることにより、 ΔP_i が算出される。

$$\begin{aligned} \Delta P_i &= P(e|i) - P(e|\neg i) \\ &= q_i + P(a|i) \cdot q_a - q_i \cdot P(a|i) \cdot q_a - P(a|\neg i) \cdot q_a \\ &= \{1 - P(a|i) \cdot q_a\} \cdot q_i + \{P(a|i) - P(a|\neg i)\} \cdot q_a \end{aligned} \quad (\text{式 6})$$

式 6 を変更することによって発生的因果効力 q_i の算出式が得られる。

(6) $P(i) = 1$ を代入して、 $P(e)$ を $P(e|i)$ に、 $P(a)$ を $P(a|i)$ に更新すること。

$$q_i = \frac{\Delta P_i - \{P(a|i) - P(a|\neg i)\} \cdot q_a}{1 - P(a|i) \cdot q_a} \quad (\text{式 7})$$

原因 a が原因 i と独立に生起する時 (i.e., $P(a|i) = P(a|\neg i) = P(a)$), 式 7 は次のように単純化される。

$$q_i = \frac{\Delta P_i}{1 - P(a) \cdot q_a} \quad (\text{式 8})$$

先行知識によって q_a が既知の場合, その知識を適用して q_i を算出する。 q_a が未知の場合, パワーアプローチが直面した因果推論の開始はいつかという問題がここでも生じる。つまり, q_i を求めるには q_a が必要となり, q_a を算出する際 q_i と同じ問題が浮上する。しかし, q_i を算出するために q_a 自体を求める必要は無く, $P(a) \cdot q_a$ が分かれば十分であり, $P(a) \cdot q_a$ は $P(e|\neg i)$ で置き換えられるので式 8 は次のようになる。

$$q_i = \frac{\Delta P_i}{1 - P(e|\neg i)} \quad (\text{式 9})$$

式 9 は発生的因果効力 q_i の一般式である。 $q_i = 1$ は原因 i が結果 e を必ず発生させることを意味し, $q_i = 0$ は原因 i と結果 e が無関係であることを意味する。この式では他の原因 a によって結果 e が常に生起している時 (i.e., $P(e|\neg i) = 1$), q_i は定義されない。対照的に, 他の原因 a によって結果が生起しない時 (i.e., $P(e|\neg i) = 0$), $q_i = \Delta P_i$ となる。つまり, 他の原因が存在しない事態において共変動と因果は一致する。また, $0 < P(e|\neg i) < 1$ の場合は $q_i > \Delta P_i$ となる。

2-3. 抑制的因果効力

抑制的因果効力の算出においても先述した仮定が用いられる。 $\Delta P_i \leq 0$ となる時, 抑制的因果効力 p_i が算出される。原因 e が生起する確率 $P(e)$ は, 原因 i が p_i で結果を抑制しない確率と原因 a が q_a で結果を発生させる確率の積により算出される。

$$P(e) = P(a) \cdot q_a \cdot \{1 - P(i) \cdot p_i\} \quad (\text{式 10})$$

原因 i が存在するという情報を付加して, 条件付き確率 $P(e|i)$ を求める。

$$P(e|i) = P(a|i) \cdot q_a \cdot (1 - p_i) \quad (\text{式 11})$$

同様に、原因 i が存在しないという情報を付加することで、条件付き確率 $P(e|\neg i)$ が算出され、式 5 と同じになる。

$$P(e|\neg i) = P(a|\neg i) \cdot q_a \quad (\text{式 5})$$

式 11 と式 5 の差分から、 ΔP_i が算出される。

$$\begin{aligned} \Delta P_i &= P(e|i) - P(e|\neg i) \\ &= P(a|i) \cdot q_a \cdot (1 - p_i) - P(a|\neg i) \cdot q_a \\ &= \{P(a|i) - P(a|\neg i)\} \cdot q_a - P(a|i) \cdot q_a \cdot p_i \end{aligned} \quad (\text{式 12})$$

式 12 を変形することによって、抑制的因果効力 p_i の算出式が得られる。

$$p_i = \frac{\{P(a|i) - P(a|\neg i)\} \cdot q_a - \Delta P_i}{P(a|i) \cdot q_a} \quad (\text{式 13})$$

原因 a が原因 i と独立に生起する時 (i.e., $P(a|i) = P(a|\neg i) = P(a)$), 式 13 は次のように書き換えられる。

$$p_i = \frac{-\Delta P_i}{P(a) \cdot q_a} \quad (\text{式 14})$$

発生的因果効力の時と同様に $P(a) \cdot q_a$ を $P(e|\neg i)$ に置き換えると、式 14 は以下ようになる。

$$p_i = \frac{-\Delta P_i}{P(e|\neg i)} \quad (\text{式 15})$$

式 15 は抑制的因果効力 p_i の一般式である。 $p_i = 0$ は原因 i と結果 e が無関係であることを意味し、 $p_i = 1$ は原因 i が結果 e を必ず抑制することを意味する。結果 e が全く生起しない時 (i.e., $P(e|\neg i) = 0$), p_i は定義されない。他の原因 a によって結果が常に生起する場合 (i.e., $P(e|\neg i) = 1$), $p_i = -\Delta P_i$ となって共変動と因果は一致する。また、 $0 < P(e|\neg i) < 1$ の場合は $p_i > -\Delta P_i$ となる。

3. 結合パワー PC 理論

現実場面では、特定の原因が単独で結果を生起させるよりも、他の原因と共

に作用することで結果を生起させることが多い。例えば、ウイルスが存在するという原因だけで病気になるという結果が生起することはあまりなく、免疫力の低下というもう一つの原因と結合することで病気になるだろう。同様に、免疫力が低下していてもウイルスが存在していなければ罹患することはあまりない。複数の原因が相互にどのような影響を及ぼして結果を生起させるかという上記のような問題に対してパワー PC 理論 (Cheng, 1997) を適用することはできない。パワー PC 理論は単一の因果効力を算出することはできても、他の事象がその因果効力にどのような影響を持つか検討できないからである。そこで、Novick & Cheng (2004) はパワー PC 理論 (Cheng, 1997) を拡張した結合パワー PC 理論 (conjunctive power PC theory) で複数の原因の相互作用を説明しようと試みた。

パワー PC 理論では原因 i と他の原因 a による 2 つの因果効力を考慮して、結果 e に対する原因 i の因果効力を算出していたが、結合パワー PC 理論では複数の原因の相互作用を計算するため、より多くの因果効力を考慮する必要が生じる。つまり、原因 i と他の原因 a (ここでは i と j 以外の原因) だけでなく、原因 i と相互に影響する原因 j 、さらに原因 i と原因 j の相互作用という 4 つの因果効力を考慮することになる。原因 i と原因 j の相互作用が結果 e に与える影響は結合因果効力 (conjunctive causal power) と呼ばれる (Novick & Cheng, 2004)。パワー PC 理論と結合パワー PC 理論の違いはこの結合因果効力の算出にある。単一の原因が持つ因果効力と類似して、結合因果効力の算出には以下の 4 つの仮定が存在する。

- (1) 原因 i と原因 j の相互作用は 2 つの原因が存在するときのみ働く。
- (2) 単一の原因や結合した原因を含め、全ての原因は独立に結果に影響を及ぼす。
- (3) 原因 a によって結果 e が発生することはあるが、抑制されることはない。
- (4) 結果 e に影響するあらゆる因果効力は、原因の生起頻度と無関係である。
- (5) 原因が無いときに結果 e が生じることはない。

結合因果効力の説明の前に、 ΔP_{ij} について説明を行う。原因 i と原因 j の

相互作用が発生する時と発生しない時の差分から ΔP_{ij} が求められる。

$$\Delta P_{ij} = P(e|i, j) - \tilde{P}(e|i, j) \quad (\text{式 16})$$

この式において $P(e|i, j)$ は原因 i と原因 j が存在する時に結果 e が発生する確率を表しており、 $\tilde{P}(e|i, j)$ は原因 i と原因 j が相互作用を起こさず独立に結果 e に影響を及ぼす期待確率⁽⁷⁾を示している。パワー PC 理論 (Cheng, 1997) において因果効力が ΔP_i の値に応じて発生的か抑制的に変化するように、結合パワー PC 理論においても ΔP_{ij} の値によって結合因果効力は発生的か抑制的に変化する。原因 i の因果効力は発生的または抑制的の 2 通りであり、原因 j の因果効力も同様に 2 通り存在し、結合因果効力に関しても 2 通り考えられるため、因果効力の算出式は 6 通り考えられる⁽⁸⁾。以下では、原因 i と原因 j の因果効力に基づいて 3 通りに場合分けを行って説明する。

3-1. 原因 i と原因 j がともに発生的因果効力を持つ場合

$\Delta P_i \geq 0$ かつ $\Delta P_j \geq 0$ である時、原因 i と原因 j はともに発生的因果効力を持つ。この時、式 16 における期待確率 $\tilde{P}_{++}(e|i, j)$ はそれぞれの因果効力を加算し、共通部分を引くことにより算出可能である。この計算式はド・モルガンの法則⁽⁹⁾を用いることで簡略化できる。つまり、原因 i または原因 j または他の原因 a によって結果 e が生じる確率は、3つの原因 (i と j と a) がともに結果を発生させない確率を 1 から引けばよい。

$$\tilde{P}_{++}(e|i, j) = 1 - (1 - q_i) \cdot (1 - q_j) \cdot \{1 - P(a|i, j) \cdot q_a\} \quad (\text{式 17})$$

ここで $\tilde{P}_{++}(e|ij)$ は原因 i と原因 j がともに発生的因果効力を持つ時の期待確率を示し、 q_i と q_j はパワー PC 理論を適用することによって次のように表される。

(7) 直接算出できず他の確率から推測するため、このような表記を用いる。

(8) 8 通りとならないのは、原因 i が発生的因果効力を持ち原因 j は抑制的因果効力を持つ場合は、両者を入れ替えても同一であるためである。

(9) ド・モルガンの法則によると、論理和の補集合は和集合を構成する集合の補集合の論理積である (e.g., $\neg(A \vee B) = \neg A \wedge \neg B$)。また、論理積の補集合は共通部分を構成する集合の補集合の論理和である (e.g., $\neg(A \wedge B) = \neg A \vee \neg B$)。

$$q_i = \frac{P(e|i, \neg j) - P(e|\neg i, \neg j)}{1 - P(e|\neg i, \neg j)} \quad (\text{式 18})$$

$$q_j = \frac{P(e|\neg i, j) - P(e|\neg i, \neg j)}{1 - P(e|\neg i, \neg j)} \quad (\text{式 19})$$

また、パワー PC 理論と同様に $P(a|i, j) \cdot q_a$ は $P(e|\neg i, \neg j)$ で置き換えられる。

$$P(a|i, j) \cdot q_a = P(e|\neg i, \neg j) \quad (\text{式 20})$$

式 18 から式 20 を式 17 に代入して解くと、以下の式が得られる。

$$\tilde{P}_{++}(e|i, j) = 1 - \frac{P(\neg e|i, \neg j) \cdot P(\neg e|\neg i, j)}{P(\neg e|\neg i, \neg j)} \quad (\text{式 21})$$

式 21 によって期待確率 $\tilde{P}_{++}(e|i, j)$ が算出可能となった。 $P(e|ij)$ と $\tilde{P}_{++}(e|i, j)$ の差分を取ることで算出される ΔP_{ij} に応じて、結合因果効力の性質は発生的あるいは抑制的となる。

$\Delta P_{++} \geq 0$ の時、式 21 に示した期待確率 $\tilde{P}_{++}(e|i, j)$ を用いて発生的結合因果効力 q_{ij} が算出される。原因 i と原因 j が存在する時に結果 e が発生する確率 $P(e|i, j)$ は、単一の原因が結果 e を発生させる確率と 2 つの原因の相互作用が結果 e を発生させる確率の和で求められる。ド・モルガンの法則によって次のように表現される。

$$P(e|i, j) = 1 - \{1 - \tilde{P}_{++}(e|i, j)\} \cdot (1 - q_{ij}) \quad (\text{式 22})$$

q_{ij} について解くことにより、発生的結合因果効力 q_{ij} の式が求まる。

$$q_{ij} = \frac{\Delta P_{++}}{1 - \tilde{P}_{++}(e|i, j)} \quad (\text{式 23})$$

$\Delta P_{++} \leq 0$ の時、式 21 に示した期待確率 $\tilde{P}_{++}(e|i, j)$ を用いて抑制的結合因果効力 p_{ij} が算出される。原因 i と原因 j が存在して結果 e が生起する確率 $P(e|i, j)$ は、単一の原因が結果を生起させる確率と抑制的結合因果効力が働かない確率の積で計算される。

$$P(e|i, j) = \tilde{P}_{++}(e|i, j) \cdot (1 - p_{ij}) \quad (\text{式 24})$$

p_{ij} について解くことにより、抑制的結合因果効力 p_{ij} を求めることができる。

$$p_{ij} = \frac{-\Delta P_{++}}{\tilde{P}_{++}(e|i, j)} \quad (\text{式 25})$$

3-2. 原因 i と原因 j がともに抑制的因果効力を持つ場合

$\Delta P_i \leq 0$ かつ $\Delta P_j \leq 0$ である時、原因 i と原因 j はともに抑制的因果効力を持つ。この場合、式 16 における期待確率 $\tilde{P}_{--}(e|i, j)$ は原因 a が q_a で結果 e を発生させる確率と原因 i と原因 j が p_i と p_j で結果 e を抑制しない確率の積で求められる。

$$\tilde{P}_{--}(e|i, j) = P(a|i, j) \cdot q_a \cdot (1-p_i) \cdot (1-p_j) \quad (\text{式 26})$$

$\tilde{P}_{--}(e|i, j)$ は原因 i と原因 j がともに抑制的因果効力を持つ時の期待確率を表している。パワー PC 理論を適用して p_i と p_j を求め、 $P(a|i, j) \cdot q_a$ を $P(e|\neg i, \neg j)$ で置き換えると、以下の式が得られる。

$$\tilde{P}_{--}(e|i, j) = \frac{P(e|i, \neg j) \cdot P(e|\neg i, j)}{P(e|\neg i, \neg j)} \quad (\text{式 27})$$

$\Delta P_{--} \geq 0$ の時、発生的結合因果効力 q_{ij} が算出される。算出方法は式 27 に示した $\tilde{P}_{--}(e|i, j)$ で ΔP_{--} を算出し、式 23 の $\tilde{P}_{++}(e|i, j)$ を $\tilde{P}_{--}(e|i, j)$ に置き換えるだけである。

$$q_{ij} = \frac{\Delta P_{--}}{1 - \tilde{P}_{--}(e|i, j)} \quad (\text{式 28})$$

$\Delta P_{--} \leq 0$ の時、式 27 に示した期待確率 $\tilde{P}_{--}(e|i, j)$ を用いて抑制的結合因果効力 p_{ij} が算出される。式 25 とほぼ同様の算出方法によって次の式が得られる。

$$p_{ij} = \frac{-\Delta P_{--}}{\tilde{P}_{--}(e|i, j)} \quad (\text{式 29})$$

3-3. 原因 i が抑制的因果効力を持ち、原因 j が発生的因果効力を持つ場合

$\Delta P_i \leq 0$ かつ $\Delta P_j \geq 0$ である時、原因 i は抑制的因果効力を持ち、原因 j は発生的因果効力を持つ。式 16 における期待確率 $\tilde{P}_{-+}(e|i, j)$ は、原因 j と原因 a が q_j と q_a で結果 e を発生させる確率と原因 i が p_i で結果 e を抑制しない確率の積で求められる。

$$\tilde{P}_{-+}(e|i, j) = q_j \cdot P(a|i, j) \cdot q_a \cdot (1-p_i) \quad (\text{式 30})$$

$\tilde{P}_{-+}(e|i, j)$ は一方が抑制的因果効力を持ち、他方が発生的因果効力を持つ時の期待確率を表している。パワー PC 理論を適用して p_i と q_j を求め、 $P(a|i, j) \cdot$

q_a を $P(e|\neg i, \neg j)$ で置き換えると、式 27 と同じ式が得られる。

$$\tilde{P}_{-+}(e|i, j) = \frac{P(e|i, \neg j) \cdot P(e|\neg i, j)}{P(e|\neg i, \neg j)} \quad (\text{式 27})$$

$\Delta P_{-+} \geq 0$ の時、式 27 に示した期待確率 $\tilde{P}_{-+}(e|i, j)$ を用いて発生的結合因果効力 q_{ij} が算出される。算出方法は他の発生的結合因果効力と同様であり、以下のように表現される。

$$q_{ij} = \frac{\Delta P_{-+}}{1 - \tilde{P}_{-+}(e|i, j)} \quad (\text{式 31})$$

$\Delta P_{-+} \leq 0$ の時、式 27 に示した期待確率 $\tilde{P}_{-+}(e|i, j)$ を用いて抑制的結合因果効力 p_{ji} が算出される。算出方法は他の抑制的結合因果効力と同様であり、次のように表される。

$$p_{ji} = \frac{-\Delta P_{-+}}{\tilde{P}_{-+}(e|i, j)} \quad (\text{式 32})$$

4. パワー PC 理論の検証

因果推論に対するパワー PC 理論の適用の妥当性に関して多くの実験や議論が為されてきた (e.g., Cheng & Novick, 2005 ; Luhmann & Ahn, 2005 ; White, 2005 ; Wu & Cheng, 1999)。Cheng (1997) は ΔP よりも因果効力の方が既存の研究結果 (e.g., Wasserman et al., 1993) を説明できると主張した。しかし、Allan (2003) は、いくつかの研究結果 (Allan & Jenkins, 1983, Experiment 3 ; Wasserman, Dorner, & Kao, 1990, Experiment 2) については十分な説明とは言えないと反論した。つまり、因果効力の方が ΔP よりも当てはまりがよいとは限らないわけである。このような論争の中で、2 つのモデルの妥当性を直接的に検証する実験が行われるようになった。 ΔP の値を一定にして因果効力の値を操作、あるいは逆に因果効力の値を一定にして ΔP の値を操作する実験事態が用いられた。このような実験事態における評定値の振る舞いからどちらのモデルが適切であるか検討されてきた。

Buehner & Cheng (1997) は、 ΔP と発生的因果効力の値について複数の条件を設定して実験を行った。 $P(e|c)$ と $P(e|\neg c)$ は 1.00, .75, .50, .25, .00

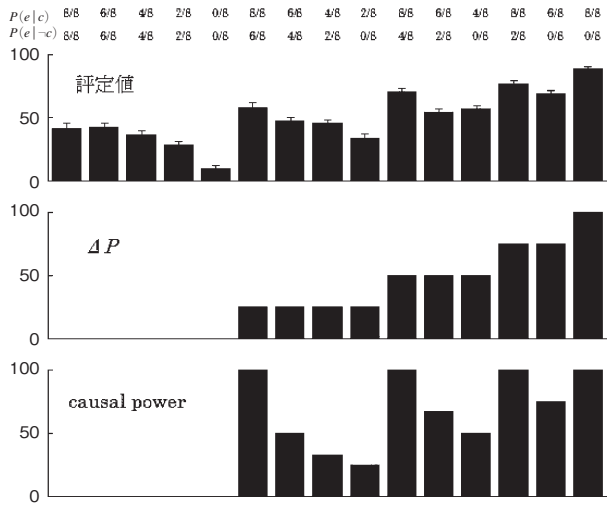


Figure 2 Buehner & Cheng (1997, Experiment 1 B) の実験結果. 上部の数値は生起頻度を, エラーバーは標準誤差をそれぞれ表す.

のいずれかの値を取り, 5 段階の ΔP (.00, .25, .50, .75, 1.00) を形成するように組み合わせられて 15 条件が設定された (Experiment 1 B)。実験参加者は 16 試行の観察とそれに続く因果関係の判断を条件ごとに行った。実験の結果, 同一の ΔP に対して発生的因果効力の減少に伴い評定値の減少がみられた (Figure 2)。しかし, $\Delta P = 0$ の条件では, いずれのモデルの予測とも一致しない結果が得られた。また, 抑制的因果効力について検討を行った実験 1 A においても同様に, 抑制的因果効力との対応と $\Delta P = 0$ における逸脱がみられた。

パワー PC 理論を支持する Buehner & Cheng (1997) の結果に異を唱えたのは Lober & Shanks (2000) であった。彼らもまた, ΔP と発生的因果効力の値について複数の条件を設定して実験を行った (Experiment 1-3)。実験参加者は 60 試行の観察を行い, 10 試行の観察ごとに因果関係の判断を行った。実験の結果, 発生的因果効力が一定であっても ΔP に対応する形で評定値が変化することが示された (Figure 3)。また, 抑制的因果効力に関して

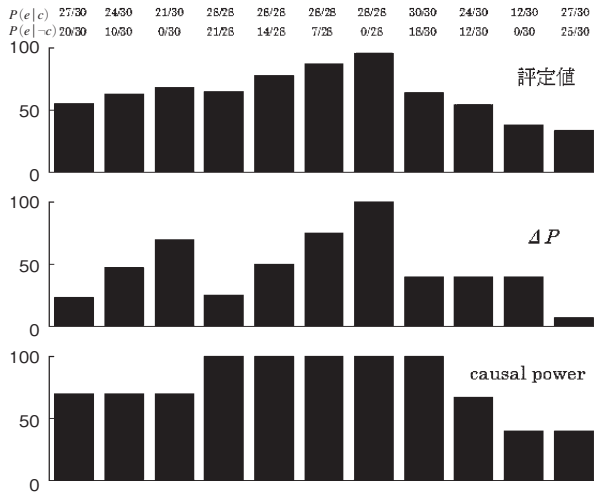


Figure 3 Lober & Shanks (2000, Experiment 1-3)
の実験結果. 上部の数値は生起頻度を表す.

もパワー PC 理論と矛盾する結果が示されている (Shanks, 2002)。同様の結果は他にも報告されている (e.g., Perales & Shanks, 2003 ; Vallee-Tourangeau, Murphy, Drew, & Baker, 1998)。

Buehner & Cheng (1997) や Lober & Shanks (2000) の結果から明らかのように、 ΔP モデルとパワー PC 理論のどちらも評定値を十分に説明できていない。評定値は ΔP に応じて変化する一方で、因果効力に応じても変化している。Buehner, Cheng, & Clifford (2003) は、判断を求める際の聞き方のあいまいな表現がこれらの結果をもたらしていると説明している。因果効力を尋ねる際には「原因がどの程度結果を引き起こすか？」という従来の質問よりも、「他の原因が結果を発生させない状況で、その原因はどの程度結果を引き起こすか？」という反実表現を用いた質問の方が適している。反事表現の質問を用いた実験ではパワー PC 理論を支持する結果が得られているが (Buehner, Cheng, & Clifford, 2003 ; Collins & Shanks, 2006), 支持しない結果も報告されている (Perales & Shanks, 2008)。

また、評定値が ΔP にも因果効力にも応じて変化する問題はサンプルサイズの問題とも考えられる。Buehner & Cheng(1997)の実験は 16 試行, Lober & Shanks (2000)の実験は 60 試行で構成されており、観察するサンプルサイズの違いが因果関係の判断に影響を与えた可能性がある。しかし、パワー PC 理論は原因の生起頻度と因果効力は無関係であるとしてサンプルサイズの要因を考慮しておらず、サンプルサイズの影響を説明することができない。Lilje-holm & Cheng (2009) はサンプルサイズの影響を説明するコンフレーション仮説 (conflation hypothesis) を提唱している。コンフレーション仮説におけるサンプルサイズは実際のサンプル数ではなく、原因の効果を受けるサンプル数である。例えば、患者 30 人中 10 人に投薬を行う場合と患者 30 人中 20 人に投薬を行う場合では、実際のサンプル数は 30 人で同じであるが、投薬の効果を受ける人数は 10 人と 20 人で異なる。この差異が因果関係の判断における確信度に影響し、評定値における差として生じてくる。

5. おわりに

本稿では、因果の認知をめぐる問題とその問題に対するいくつかのアプローチについて論じてきた。はじめに、連合的アプローチや法則基礎的アプローチに代表される共変動アプローチとそれに対立するパワーアプローチという異なる 2 つの立場を説明した。そして、2 つのアプローチにおける問題点を克服するための新たな立場である因果的アプローチとして、両者の統合を試みたパワー PC 理論 (Cheng, 1997) について概説した。また、パワー PC 理論を拡張した結合パワー PC 理論 (Novick & Cheng, 2004) について紹介し、拡張することで説明可能になった複数の原因事象の相互作用について概略を述べた。さらに、パワー PC 理論の妥当性を検証した実験 (e.g., Buehner & Cheng, 1997; Lober & Shanks, 2000) を展望した。

因果推論の計算モデルとしてパワー PC 理論が妥当であるか否かについてはっきりとした結論は得られていないが、このモデルが因果推論の研究に大き

く貢献したことは間違いない。パワー PC 理論の妥当性を検証する中で、新たに多くのモデルが考案されてきた。法則基礎的アプローチとして、服部 (2001) は二要因ヒューリスティックス・モデル (dual-factor heuristics model) を、Perales & Shanks (2007) は EI ルール (evidence integration rule) を、White (2008) は加重平均モデル (weighted averaging model) をそれぞれ提唱している。因果的アプローチでは Gopnik, Glymour, Sobel, Schulz, Kushnir, & Danks (2004) が因果マップ (causal maps) を、Griffiths & Tenenbaum (2005) が因果サポートモデル (causal support model) を提唱した。また、パワーアプローチからは力学モデル (dynamics model) が Wolff (2007) によって発表されている。

パワー PC 理論はその後、因果ベイジアンネットワーク (causal Bayesian network) の特殊なタイプと同値となることが証明されている (Glymour, 2001, 2003)。因果ベイジアンネットワークは複数の事象の因果関係を非循環有向グラフで、その強さを条件付き確率で表す数学的手法である (Pearl, 2000 ; see Gopnik, Glymour, Sobel, Schulz, Kushnir, & Danks, 2004 for a general proposal of such networks as a psychological model)。パワー PC 理論は複数の原因と 1 つの結果という事態にしか適用できなかったが、因果ベイジアンネットワークでは複数の原因と複数の結果を表現することができる。パワー PC 理論は、因果ベイジアンネットワークを用いる SS パワーモデル (sparse and strong power model : Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008) へと更なる発展を遂げている。SS パワーモデルについては稿を改めて論じたい。

引用文献

- Allan, L. G. (2003). Assessing power PC. *Learning & Behavior*, 31, 192–204.
- Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning and Motivation*, 14, 381–405.
- Buehner, M. J., & Cheng, P. W. (1997). Causal induction : The Power PC theory versus the RW model. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the*

- Nineteenth Annual Conference of the Cognitive Science Society* (pp. 55–69). Hillsdale, NJ: Erlbaum.
- Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1119–1140.
- Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, 18, 537–545.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367–405.
- Cheng, P. W., & Novick, L. R. (2005). Constraints and Nonconstraints in Causal Learning: Reply to White (2005) and to Luhmann and Ahn (2005). *Psychological Review*, 112, 694–707.
- Collins, D. J., & Shanks, D. R. (2006). Conformity to the Power PC theory of causal induction depends on the type of probe question. *Quarterly Journal of Experimental Psychology*, 59, 225–232.
- Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, 47, 109–121.
- Denniston, J. C., Savastano, H. I., & Miller, R. R. (2001). The extended comparator hypothesis: Learning by contiguity, responding by relative strength. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 65–117). Mahwah, NJ: Erlbaum.
- Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.
- Glymour, C. (2003). Learning, prediction, and causal Bayes nets. *Trends in Cognitive Sciences*, 7, 43–48.
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, 111, 1–30.
- Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, 51, 334–384.
- 服部雅史 (2001). 因果帰納の二要因ヒューリスティックス・モデル. *認知科学*, 8, 444–453.
- Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. *Cognitive Science*, 31, 765–814.
- Jenkins, H., & Ward, W. (1965). Judgment of contingency between responses

- and outcomes. *Psychological Monographs*, 7, 1–17.
- Liljeholm, M., & Cheng, P. W. (2009). The influence of virtual sample size on confidence and causal strength judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 157–172.
- Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review*, 107, 195–212.
- Lu, H., Yuille, A., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, 115, 955–984.
- Luhmann, C., & Ahn, W.-k. (2005). The meaning and computation of causal power: Comment on Cheng (1997) and Novick and Cheng (2004). *Psychological Review*, 112, 685–693.
- Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 51–92). San Diego, CA: Academic Press.
- Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review*, 111, 455–485.
- Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, 94, 61–73.
- Pearl, J. (2000). *Causality*. Cambridge, England: Cambridge University Press.
- Penn, D. C., & Povinelli, D. J. (2007). Causal cognition in human and nonhuman animals: A comparative, critical review. *Annual Review of Psychology*, 58, 97–118.
- Perales, J. C., & Shanks, D. R. (2003). Normative and descriptive accounts of the influence of power and contingency on causal judgment. *Quarterly Journal of Experimental Psychology*, 56 A, 977–1007.
- Perales, J. C., & Shanks, D. R. (2007). Formal models of causal judgment: A review and synthesis. *Psychonomic Bulletin & Review*, 14, 577–596.
- Perales, J. C., & Shanks, D. R. (2008). Driven by power? Probe question and presentation format effects on causal judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1482–1494.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
- Sawa, K. (2009). Predictive behavior and causal learning in animals and hu-

- mans. *Japanese Psychological Research*, 51, 222–233.
- Shanks, D. R. (2002). Tests of the power PC theory of causal induction with negative contingencies. *Experimental Psychology*, 49, 81–88.
- Shanks, D. R. (2007). Associationism and cognition : Human contingency learning at 25. *Quarterly Journal of Experimental Psychology*, 60, 291–309.
- 嶋崎恒雄 (1995). 共変動査定事態の情報処理に関する 2 つのモデル. 人文論究 (関西学院大学), 45 (1), 17–31.
- Vallée-Tourangeau, F., Murphy, R. A., Drew, S., & Baker, A. G. (1998). Judging the importance of constant and variable candidate causes : A test of the power PC theory. *Quarterly Journal of Experimental Psychology*, 51 A, 65–84.
- Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak & D. L. Medin (Eds.), *The psychology of learning and motivation*, Vol. 34 : *Causal learning* (pp. 47–88). San Diego, CA : Academic Press.
- Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength : The role of structural knowledge and processing effort. *Cognition*, 82, 27–58.
- Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given : Causal models in learning and reasoning. *Current Directions in Psychological Science*, 15, 307–311.
- Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models : Asymmetries in cue competition. *Journal of Experimental Psychology : General*, 121, 222–236.
- Waldmann, M. R., & Martignon, L. (1998). A Bayesian network model of causal learning. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 1102–1107). Mahwah, NJ : Erlbaum.
- Wasserman, E. A., Dorner, W. W., & Kao, S. F. (1990). The contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology : Learning, Memory, & Cognition*, 16, 509–521.
- Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations : Role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology : Learning, Memory, & Cognition*, 19, 174–188.
- White, P. A. (2005). The power PC theory and causal powers : Comment on Cheng (1997) and Novick and Cheng (2004). *Psychological Review*, 112, 675

-684.

White, P. A. (2008). Accounting for occurrences : A new view of the use of contingency information in causal judgment. *Journal of Experimental Psychology : Learning, Memory, & Cognition*, 34, 204-218.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology : General*, 136, 82-111.

Wu, M., & Cheng, P. W. (1999). Why causation need not follow from statistical association : Boundary conditions for the evaluation of generative and preventive causal powers. *Psychological Science*, 10, 92-97.

——斎藤元幸 大学院文学研究科博士課程前期課程——

——嶋崎恒雄 文学部教授——