The 22nd Annual Conference of the Japanese Society for Artificial Intelligence, 2008

1M2-5

因果推論とアドホック論理：世界内に留まり続ける意識の構造

郡司ペギオ幸夫[*1]
Yukio-Pegio Gunji

[*1] 神戸大学理学研究科地球惑星システム科学専攻
Department of Earth & Planetary Sciences, Faculty of Science, Kobe University, Nada Kobe 657-8501, JAPAN

1. はじめに

部屋を片付けないとおもちゃを買ってあげません。そういわれた子供は、夢中で部屋を片付けるだろう。片付ければおもちゃが買ってもらえると信じるからだ。「部屋を片付けない」を A,「おもちゃを買わない」を B とすると、親の発言が、$A\rightarrow B$ である一方、子供の期待は $\neg A\rightarrow\neg B$ と表される。子供の期待の対偶は $B\rightarrow A$ であるから、子供はむしろ $A\rightarrow B$ と $B\rightarrow A$ を同一視しているかのようだ。このような因果関係における原因と結果の混同を、因果推論における対称性バイアスと呼ぶ。

この対称性はアリストテレス以来度々取り上げられてきた問題であるが、それがどの程度人間の認知に寄与しているか、実験的に定量化されたのはごく最近のことだ。特に 2000 年以降、原因と結果に関する特定の組み合わせを被験者に与え、因果関係を推定させる実験が標準化されて、認知的誤謬の構造が明らかになりつつある。

ではこのような認知的誤謬の起源はどこにあるか。因果関係の対称性は、現実世界の対象と、その表象、つまりモノと言葉の関係においても論じられる。モノを指し示す言葉の選択を学習することは、条件付けによって多くの動物で学習可能であるが、その言葉の指示するモノ選択の学習はほとんどの動物で不可能だ。つまり言葉からモノ、モノから言葉という対応を対称的に自由に行えるのは人間に固有で、これもまた認知的対称性と呼ばれている。

本論文は、モノと言葉の非対称性が、因果関係における対称性を創りだしているというモデルを提案する。モノの世界を n ビット列をすべて集めた世界（集合束）によって、それを表象する言葉を、集合束の部分集合を取り出し、論理として完備化した最大元交構造で表し、全体・部分という覆しがたい非対称性を対称化しようとする運動が、認知活動に潜み、これが絶えず構成される論理（アドホック論理）に論理的誤謬を与えることを示す。またこのモデルが標準化された実験結果のほとんどを説明することを示す。

2. 集合束と最大元交構造

2.1 トークン世界としての集合束

モノ世界と言葉の世界との相互作用・調停をモデル化するため、ここでは世界における対象（トークン）をビット列で表す。世界には構造があって、これは順序関係でモデル化される。モノの間、つまりビット列間で、次のようなビット列順序を考える：二つのビット列 $a_1a_2\cdots a_n$ および $b_1b_2\cdots b_n$ において、任意の i に関して $a_i\leq b_i$ であるとき、ビット列 $a_1a_2\cdots a_n$ よりビット列 $b_1b_2\cdots b_n$ が大きいという。もし $a_1a_2\cdots a_n$ を A、$b_1b_2\cdots b_n$ を B と書くなら、ビット列の大小関係を $A\leq B$ と書く。ビット列(特定の n ビットのみ)の適当な集合を考え、これを S とする。いま S の部分集合 X を考え、X のどのビット列よりも大きいビット列を X の上界という。上界の中で最小の（ビット列の大小関係の意味で）ビット列を X の上限という。下界、下限も双対的に定義される。ここで、$X=\{A, B\}$なる二つのビット列からなる集合とすると、その上限、下限を、$A\cup B$（交と呼ぶ）、$A\cap B$（和と呼ぶ）と表すとき、それらは以下のように具体的に書き下せる。

$$A\cup B = \max(a_1, b_1)\max(a_2, b_2)\cdots\max(a_n, b_n)$$
$$A\cap B = \min(a_1, b_1)\min(a_2, b_2)\cdots\min(a_n, b_n).$$

すべてのビット列から構成された順序集合は、交、和について閉じており集合束をなす。これをトークン世界のモデルと考える。

2.2 タイプ世界としての最大元交構造

世界の中に生きる者が鍛え上げ、獲得する言葉は、トークンの集合を一般化し、タイプ化したものと考えられる。ここでは、これを集合束の部分と考える。ただし部分束ではなく、集合束の部分集合 S から最大元交構造として構成されたものと考える。S の任意の二つのビット列 A と B について $A\cap B$ を計算し、$A\cap B$ が S に属していないなら、これを新たに S に加える。こうして $A\cap B$ によって新たなビット列が現れなくなるまで、この操作を繰り返す。最後に n ビット列における最大の元である、全ての桁が1のビット列をこれに加える。以上の操作をまとめて、最大元交操作と呼び、T で表し、得られる集合 $T(S)$を最大元交構造という。

最大元交構造 $T(S)$におけるビット列 A と B の上限、下限を、$A\vee B$, $A\wedge B$ で表すとき

$$A\vee B = \cap\{X\in T(S) \mid A\leq X,\ B\leq X\}$$
$$A\wedge B = A\cap B$$

と表せる。下限の定義は集合束と同じであるが、上限が異なる。

この $T(S)$が絶えず、外部と相互作用して論理構造を維持しながら、外部と内部、トークンとタイプの対称化を進めるとき、束は特定の構造をとりやすくなる。外部との相互作用を実装した論理を、アドホック論理と呼ぶ。

3. アドホック論理

ここでタイプ世界は、世界に対して十分小さく計算資源を節約した論理であると仮定する。しかしそれは、絶えず論理外部と

連絡先：郡司ペギオ幸夫，神戸大学理学研究科，657-8501 神戸市灘区六甲台１−１，078-803-5759，yukio@kobe-u.ac.jp

相互作用し、外部にあるビット列を新しい情報として、自らのうちに取り込もうとする。情報の取り込みと、あまり束を大きくしないとの相反する要請を満たすため、新奇な情報の取り込みには以下のような刈り込み操作を伴うことになる。

3.1 刈り込み操作

アドホック論理では、新たなビット列 C が付加されるとき、既存の束 L に次のような事前刈り込み操作を施す:

$$F(L) = L - \{X \in L \mid X \cap C \notin L^*\}.$$

ここで $P-Q$ は集合 P から集合 Q を取り除く操作を意味する。$L^* = L-\{C\}$であり、ビット列 C が真に新しいビット列で、$C \notin L$ である限り $L^* = L$ である。束 L に施される操作 F は、新たなビット列 C を付加する前に、C と相互作用して過剰に新たなビット列を生産するビット列 X を取り除く操作となっている。$L^* = L-\{C\}$ によって、客観的に新たなビット列でないものに対しても事前刈り込みが可能となる。

3.2 アドホック論理の時間発展

まず、9 ビット列の集合 S から最大元交操作で束 $T(S)$が得られる。ここで $A \rightarrow B = A^c \vee B = 1$（最大元）を満たす 2 つのビット列 A および B が選ばれ、続いて $B \rightarrow A$ を計算する。このとき B^c が束内に存在しないなら、新たに外部から条件を満たし、束に加える。ただし A^c は $A^c \vee A = 1$かつ $A^c \wedge A = 0$（アドホック論理の最小元）を満たす A の補元。とくに B^c が存在する場合でも、確率 P でそれは新たな元と認識され、刈り込み操作が適用される。その後、最大元交操作で束となる。これを繰り返すことで、時間発展を実現する。

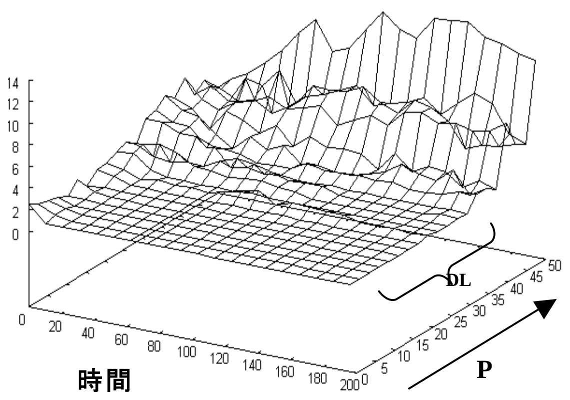

図1．アドホック論理の時間発展．図中の P は発見された束内の元を外部にある新たな元とみる確率（x100）で、縦軸は1個の元あたりの平均補元数で、1なら分配束．それ以上なら非分配束．

P が 0.4 以上のとき非分配相補束となるが（図1）、このときこのような束において $A \rightarrow B = A^c \vee B = 1$ が成立するとき、$B \rightarrow A = 1$ の成立する確率も高くなり（ほぼ 0．8）、対称性バイアスが獲得されているといえる。

3.3 因果推論

アドホック論理が、因果推論に関する実験データをどの程度説明するか評価してみる。実験は次のような形式をとる。被験者の目の前にモニターがあし、それは左右に分けられ、左に原因となる画像、右に結果となる画像が映され、常に原因・結果の対として提示される。原因は一方を A とすると他方はその否定、結果は一方を B とすると他方はその否定があり、各々2 種類ある。つまり被験者は 4 種類の組み合わせを特定の確率分布に従ってみせられる。この分布が実験条件となり、特定の条件下モニターを見た後、$A \rightarrow B$ を真と思うか、その確信度を問われる。こうして因果対の結合確率分布に対し、確信度が得られる。

我々は因果対の結合確率分布が与えられるとき、これを以下のような方法でアドホック論理に埋め込んだ（否定は補元とする）。(1)ランダムに 7 個の 7 ビット列を選び、T で束とし、(2)因果対の結合確率が相対的に高いときには、結合した原因と結果が順序関係を持ち、低いときには持たないとして(1)で得た束に条件を埋め込み（ただし$\neg A$ と$\neg B$ との結合は無視する）、(3)結合確率 $P(A, B)/P(\neg A, \neg B)>1$ のとき元 A, B および $\neg A, \neg B$ に刈り込み操作を適用する。こうして得られた束で $A \rightarrow B == 1$となるか否かを判定し、これを 1000 回繰り返して因果律成立確率を求めた。

以下に文献にある実験データの確信度とアドホック論理における因果律成立確率間の相関係数を示す。これ以外の実験データについても検討したが、いずれも同様に高い相関を示した。

AS95	WDK90	BC3.1	BC3.3	HO7.1	HO7.2	LS00
0.978	0.824	0.915	0.902	0.814	0.984	0.835

ここでも刈り込み操作が重要な意味を持っている。

4. 結論

因果関係に関する認知的誤謬の起源が、トークン世界の一部を凡化してタイプとする操作に求められるという作業仮説をもって、これを集合束と、その部分集合をとって完備化し、外部と相互作用するアドホック論理によってモデル化した。このモデルでは、内と外の非対称性を絶えず対称化しようとする運動が、論理内部にねじれた対称性を作り出すと考えられ、これがほどほどの対称性バイアスを作り出すこと、因果推論に関する実験データをよく説明することが示された。

参考文献

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(6), 1119-1140.

郡司ペギオ幸夫・澤宏司(2008)認知的誤謬の起源：アドホック論理と対称性バイアス. 認知科学（投稿中）.

Hattori, M. & Oaksford, M. (2007). Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. Cognitive Science, 31(5), 765-814.

Lober, K. & Shanks, D.R. (2000). Is causal induction based on causal power? Crtique of Cheng (1997). Physical Revies 107(1), 195-212

White, P.A. (2003).Making causal judgments from the proportion of confirming instances: The *pCI* Rule. Journal of Experimental Psychology: Learning, Memory and Cognition 29(4), 710-727.