

仮説検証過程における確信度更新と検証系列： 情報獲得モデルによる検討

田村 昌彦・服部 雅史・三輪 和久

In this paper, we propose a new model of human hypothesis testing process. The model predicts a sequence of testing when people continuously receive positive results that confirm the hypothesis. First we constructed a Bayesian model which addresses 5 situations to be tested based on information theory. In a simulation, we identified 5 qualitative characteristics of the model's behavior, some of which were corroborated in a subsequent experiment. The results of the experiment, however, also suggested that the participants did not correctly estimate the size of the hypothesis. By reconstructing the model, we simplified the model to explain participants' data. As a consequence, we found that a much simpler model can sufficiently explain the data. In conclusion, it was suggested that people are roughly Bayesian but they use heuristic strategies in hypothesis testing.

Keywords: Wason's 2-4-6 task (Wason の 2-4-6 課題), expected information gain (期待獲得情報量), positive test strategy (正事例検証方略), Bayesian model (ベイズ・モデル)

現実世界は曖昧さに満ち溢れているが、われわれは有限の情報処理能力しか持たない。したがって、われわれの知識は厳密にはすべて仮説的である。仮説の正しさが論理的に証明できる場合はむしろ稀で、実際には、現実の事例から仮説の妥当性を判断することになる。事例から仮説の正しさを評価する際、典型的には2つの方法がある。1つは、正事例検証 (positive test; PT) と呼ばれるもので、仮説に合致する事例が本当に正しいか確かめる方法である。もう1つは、負事例検証 (negative test; NT) と呼ばれるもので、仮説に合致しない事例が本当に間違っているか確かめる方法である。

Wason (1960) が提案した 2-4-6 課題が、このような仮説検証過程を調べるための課題としてよく知られている。この課題では、はじめに問題解決者 (以下、解決者) に「2-4-6」という「3つ組」の整数

と、これらがある規則 (正解の規則) に則って構成されている旨を伝えられる。解決者は、この規則を予測し、具体的な事例を生成して検証することを求められる。実験者は、解決者が生成した事例が規則に則っているか否かをフィードバックする。この手続きを解決者が正解の規則を発見するまで続ける。多くの解決者は、はじめに「偶数列」や「2ずつ増える数」などの仮説を形成する。その後、「4-6-8」や「10-12-14」などの PT を多用して仮説を支持する結果を受けるため、この仮説が正しいと思ってしまう。しかし、この課題の正解は、「上昇する数列」である。仮説が誤っていることに気づくためには、「1-2-15」などの NT が有効になるが、そのことに気づかず、なかなか正解にたどり着けない。

このように、仮説検証の際、われわれは、自分の仮説に合う事例を選択的に認知したり、重視したりする傾向があることが知られている。この傾向は、確認バイアス、または正事例検証方略と呼ばれている。ただ、「確認」とは、検証の方法や傾向ではなく検証の結果を指すことばであり、混乱を招きや

すいため、本論文では、Klayman & Ha (1987) にならって正事例検証方略（以下では、PT 方略）と呼ぶ。

PT 方略は、正しい帰納を妨げるバイアスで、誤信念やステレオタイプの認知などの、望ましくない社会的認知の一因とも考えられてきた（たとえば、Nisbett & Ross, 1980）。しかし、近年になって、PT の適応的合理性が指摘されている（たとえば、Friedrich, 1993; Klayman & Ha, 1987）。PT の結果が肯定的であったとき、仮説が正しいためにその結果が得られる場合と、全くの偶然でそうなる場合が考えられるが、後者の確率が十分に低いことが期待できる環境においては、PT は、仮説を効率的に検証するための方略となる（Klayman & Ha, 1987）。Hoenkamp (1989) は情報理論を用いて、最適な検証の選択は最も曖昧さを低減できる選択であることに注目した。彼は 2-4-6 課題における NT の有用性や、PT 方略が合理的な方略である指摘（たとえば、Gorman & Gorman, 1984; Klayman & Ha, 1987）がこの枠組みで説明できるとした。また、情報獲得理論（Oaksford & Chater, 1994）によれば、Wason 選択課題（Wason, 1966）において最もよく見られる「非論理的な」回答は、仮説検証の効率性において最も合理的な PT 選択である。

これまでの仮説検証に関する研究は、仮説の修正のために世界からいかに有効な情報を引き出すかという点に重きが置かれてきた。仮説の修正に有効な情報とは、より具体的に言えば、仮説に対する反証情報である。正しい仮説の発見や効率的な検証のためには、反証情報が重要となる。反証情報が隠蔽されることによって、われわれはしばしば誤った帰納に陥ることになる。つまり、これまでの議論は、世界から反証情報を引き出すという意味において、PT は合理的な仮説検証の方法であるということを示したものである。

一方で、仮説検証の最終段階では、われわれは世界から反証情報を受け取ることがなくなる。さらに言えば、最終的に仮説が世界の真実に到達した時点では、検証の結果は常に仮説を支持する状況になる。これは、上記の観点から言えば、世界から何ら有効な情報を得ることができない状況である。仮説の正しさは永遠に蓋然的である。したがって、世界から仮説修正に関して有効な情報を得ることができない状況においては、その仮説の蓋然性をもっとも

増加させる仮説検証の方法が重要になる。このような観点に立った仮説検証の研究は、著者らが知る範囲で行われてきていない。つまり、仮説検証過程の初期段階から最終段階にかけて、どのように検証のタイプを選択していくのかという点について、検証系列をモデル化した研究は見あたらない。そこで、本研究では、このような状況における仮説検証過程の方略の遷移を、Bayes 的モデルを用いて分析し、モデルの予測を人間を対象とした実験結果と比較することにより検討した。

1. モデルの構築

帰納的推論では、ある仮説に対して、どれだけ検証をくり返し、その仮説が支持されても、正解に到達する保証はない。しかし、われわれは自身の持つ仮説が支持され続けることで、その仮説が正しいと判断するだろう。本研究では、このように常に仮説が支持され続けた場合の仮説検証過程を扱う。

ここでは、形成された仮説に対して、初期のあまり確信のない状態から、最終的に強い確信をいだくようになる状態までの過程を、Bayes 的な仮説更新の考えに基づいてモデル化した。前述のとおり、事例によって規則を検証する方法には 2 つある。本研究でもこれら 2 つを扱った。

1.1 モデルで扱う状況

解決者が想定している仮説と正解の規則の間の論理的な関係がどうなっているか、その可能性について考える必要がある。仮説に合う事例の集合を H 、正解の規則に合う事例の集合を T とすると、 H と T およびそれらの補集合 $\bar{H}$, $\bar{T}$ によって規定される部分集合は、 $H \cap T$, $H \cap \bar{T}$, $\bar{H} \cap T$, および $\bar{H} \cap \bar{T}$ の 4 つとなる。これら 4 つの部分集合が空集合か否かによって $2^4 = 16$ 通りの状況を考えることができる。しかし、いくつかの状況は、一般的な仮説検証状況としては不自然である。たとえば、 H が空集合だとすると、仮説に合う事例が 1 つもないことになるが、通常はそのような仮説を立てることはない。 T が空集合である状況は、正解となる規則に合う事例が 1 つもない状況であるが、そのような規則を問題にするような事態は考えにくい。同様に、 $\bar{H}$ や $\bar{T}$ が空集合である状況も不自然である。これらはどのような事例でもよいという仮説（ $\bar{H}$ が空集合）や規則（ $\bar{T}$ が空集合）を示す。よって、 H , T ,

表1 仮説 H と正解規則 T の論理的関係

	S_3	—	S_2	—	S_1	—	S_0	—	S_4	—	—	—	—	—	—
$H \cap T$	$\exists$	$\exists$	$\exists$	$\exists$	$\exists$	$\exists$	$\exists$	$\exists$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
$H \cap \bar{T}$	$\exists$	$\exists$	$\exists$	$\exists$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\exists$	$\exists$	$\exists$	$\exists$	$\emptyset$	$\emptyset$	$\emptyset$
$\bar{H} \cap T$	$\exists$	$\exists$	$\emptyset$	$\emptyset$	$\exists$	$\exists$	$\emptyset$	$\emptyset$	$\exists$	$\exists$	$\emptyset$	$\emptyset$	$\exists$	$\exists$	$\emptyset$
$\bar{H} \cap \bar{T}$	$\exists$	$\emptyset$	$\exists$	$\emptyset$	$\exists$	$\emptyset$	$\exists$	$\emptyset$	$\exists$	$\emptyset$	$\exists$	$\emptyset$	$\exists$	$\emptyset$	$\emptyset$
$H = \emptyset$													$\checkmark$	$\checkmark$	$\checkmark$
$T = \emptyset$													$\checkmark$	$\checkmark$	$\checkmark$
$\bar{H} = \emptyset$				$\checkmark$				$\checkmark$				$\checkmark$			$\checkmark$
$\bar{T} = \emptyset$						$\checkmark$		$\checkmark$					$\checkmark$		$\checkmark$
$\bar{H} \cap \bar{T} = \emptyset$		$\checkmark$		$\checkmark$		$\checkmark$		$\checkmark$		$\checkmark$		$\checkmark$		$\checkmark$	$\checkmark$

注) 表の上半分は、各部分集合に1つでも要素が存在するか($\exists$)、空集合($\emptyset$)かを示す。下半分は、 H , T , $\bar{H}$, $\bar{T}$, $\bar{H} \cap \bar{T}$ がそれぞれ空集合かどうかを示す。

$\bar{H}$, $\bar{T}$ のうち少なくとも1つが空集合の状況はモデルから除外した。さらに、これらのうち、 $\bar{H} \cap \bar{T}$ が空集合の場合もやはり不自然な状況である。これは、仮説と未知の規則は一致しないが、仮説と規則だけで全事例を網羅している(と解決者が想定している)という状況である。このような状況はあり得ないわけではないが、極めて稀な状況で、現実にはまず起こり得ないと考えるのが妥当である。よって、モデルから除外した。表1に4つの部分集合が空集合か否かによって規定される16通りの組み合わせと(上部)、除外理由(下部)を示した。最終的に、図1に示した5つの状況が残った。これらのうち、 S_0 だけが「本当に仮説が正しい」状況であり、検証過程においては、この状況に対する確信度が上昇することが期待される。

1.2 仮説と正解規則の関係

仮説を満たす事例の集合 H の大きさを $P(H)$ で表すこととする。これは、全事例の中から1つの事例を無作為抽出したときに、その事例が仮説に合う確率である。同様にして、正解規則の事例の集合 T について $P(T)$ を定める。 $P(H)$ と $P(T)$ の関係を考えるにあたり、明確にしておくべきことがある。それは、このモデルが扱うのは参加者(あるいは解決者)の内的過程だという点である。このことは、集合 T と図1の5つの状態の扱い方を決める。実験者にとっては、集合 T は固定した不変のものであり、5つの状況も(仮説がわかれば)どれが実現しているのかを1つに同定できる。しかし、集合 T は、解決者が想定する正解規則に対応し、5つの状況も解決者が想定するものであると考えれば、 $P(H)$ が状況によって異なることはあり得ないので、 $P(T)$ の方が状況 (H との関

係) 次第で変わると考えるのが妥当である。ただし、 H は T に一致することが期待されている以上、 $P(T)$ はできるだけ $P(H)$ に近くなるよう設定するのが妥当と考えた。そこで、 S_0 はもとより、 S_3 でも $P(H) = P(T)$ とした。 S_4 においては、 $P(H) = P(T)$ とすると、 $P(H) \geq .50$ のときに不都合が生じるため、 $P(H)$ が小さいほど $P(H)$ と $P(T)$ が近似するよう $P(T) = P(H)(1 - P(H))$ とした。

S_1 と S_2 における $P(H)$ と $P(T)$ の関係を規定するため、次のような被覆度と呼ぶパラメータ c ($0 \leq c \leq 1$) を導入した。

$$c = \frac{P(T|S_2)}{P(H)} = \frac{1 - P(T|S_1)}{1 - P(H)} \quad (1)$$

これは、 S_2 においては、 T が H のどれくらいの部分を被覆しているかを表す数値である。被覆度は、 S_2 において PT を行ったとき、仮説を支持する結果(“YES”)が返ってくる確率に等しくなる。同様にして、 S_1 における $P(H)$ と $P(T)$ の関係は、 S_1 において NT を行ったとき、仮説を支持する結果(“NO”)が返ってくる確率によって規定されるものとした。

1.3 確信度とその更新

図1の5つの状況 S_i に対して、 $\sum_i P(S_i) = 1$ となるようにそれぞれ信念 $P(S_i)$ を割り当てて、これを状況 S_i の確信度と呼ぶ。また、 S_0 は参加者の仮説が正解と一致する状況なので、 $P(S_0)$ を特に仮説の確信度と呼ぶ。

各状況の確信度は、Bayes の定理に従って更新される。仮説を支持する結果(positive feedback; FB_P)が返る確率を $P(FB_P)$ とすると、状況 S_i

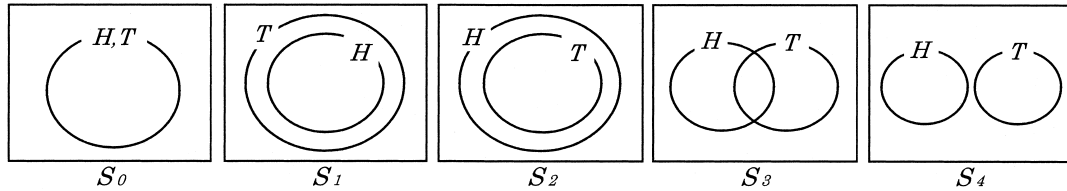


図1 モデルで扱う5つの状況. 仮説 H と正解規則 T の関係が網羅的に示されている.

の事後確信度は、次の式であらわすことができる.

$$P(S_i|FB_P) = \frac{P(FB_P|S_i)P(S_i)}{P(FB_P)} \quad (2)$$

ここで、 $P(FB_P|S_i)$ は尤度と呼ばれ、状況 S_i で仮説を支持する結果が返る確率である. この値は、行う検証が PT か NT によって異なる. 尤度の一覧を表2に示した. 尤度の算出方法は付録Aに示した.

以上の仮定に基づき、PT または NT を逐次的に使用して、式(2)にしたがって状況の確信度を更新していくわけであるが、どちらの検証を用いるかは、Shannon エントロピー（以下、単にエントロピー）に基づいて決定した. 状況 S のエントロピーは次の式で定義される.

$$H(S) = - \sum_{i=0}^4 P(S_i) \log_2 P(S_i) \quad (3)$$

仮説を支持する結果を得た後の各状況の確信度 $P(S_i)$ から、PT を用いる場合と NT を用いる場合のそれぞれのエントロピーを算出し、その値が小さい方の検証を常に選択することとした¹⁾. ただし、両方のエントロピーが等しい場合には PT を選択することとした.

2. シミュレーション

仮説の大きさ $P(H)$ と被覆度 c をモデルのパラメータとみなして、 $P(S_n)$ の変化と検証タイプの推移をコンピュータ上でシミュレートした. 無差別の原理（たとえば、Fine, 1973）から、各状況の初期確信度は一様分布とした ($P(S_i) = .20$). 図2にシミュレーション結果の例 ($P(H) = .50$, $c = .50$) を示した.

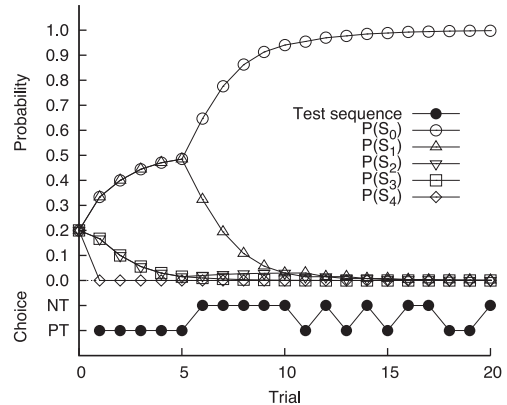


図2 シミュレーション結果の例 ($P(H) = .50$, $c = .50$ の場合).

2.1 モデル出力の定性的特徴

シミュレーション結果から、以下に挙げる5つの定性的な特徴がみられた.

- (1) $P(H)$ が小さい場合には PT から開始し、しばらく PT が続く PT フェーズの後に、NT フェーズへ移行する. ($P(H)$ が一定以上、大きい場合には NT から開始し、NT フェーズ後、PT フェーズへ移行する.)
- (2) $P(H)$ の大きさに関わらず、多くの局面で PT から検証を開始することが有効であり、NT から開始されるのは $P(H) = .83$ 付近より大きいときである.
- (3) PT フェーズから NT フェーズ、あるいは NT フェーズから PT フェーズへ移行する際に、仮説の確信度が急激に上昇する.
- (4) 仮説の確信度が十分に高まった後、PT と NT が頻繁に入れ替わる.
- (5) $P(H)$ が状況確信度の推移に与える影響は意外に小さい.

特徴 (1) は $P(H)$ が小さい場合には PT が、 $P(H)$ が大きい場合には NT が有効であるという

1) 仮説が複数存在する場合には、両者を組み合わせた診断検証 (diagnostic test; DT) と呼ばれる方法もあるが、本研究では、最も基本的な PT と NT のみを扱う.

表2 検証によって仮説を支持する結果が得られる尤度 $P(FB_P|S_i)$

検証タイプ	$P(FB_P S_0)$	$P(FB_P S_1)$	$P(FB_P S_2)$	$P(FB_P S_3)$	$P(FB_P S_4)$
PT	1	1	c	$P(H)$	0
NT	1	c	1	$1 - P(H)$	$1 - P(H)$

注) c は被覆度をあらわす。詳しくは本文を参照。

主張 (Klayman & Ha, 1987) に一致する。しかし、特徴 (2) は検証タイプの有効性が対称的ではないことを示した。つまり、 $P(H) > .50$ でも PT を用いて検証を開始する。モデルでは、PT と NT の反転が確認されたのは、 $P(H) = .86$ 付近である ($c = .50$ のとき)。このことは、多くの局面で PT が有効な検証タイプであることを示している。図 3 に $P(H)$ と c をそれぞれ .01 ずつ変動させたときの初回検証タイプを示した。図 3 から明らかなように、これら特徴 (1) の前半および (2) は、 c の値に依存しない特徴であることもわかった。PT を行った後のエントロピーを $H(S|T_P)$ 、NT を行った後のエントロピーを $H(S|T_N)$ とすると、 $P(H)$ が大きくなるにしたがって、 $H(S|T_P)$ は増加するのに対して、 $H(S|T_N)$ は減少する。これは PT の有用性が $P(H)$ の増加に伴って減少し、NT の有用性が増加することを示す。しかし、 c がどのような値でも、 $H(S|T_P)$ が $H(S|T_N)$ より上回る $P(H)$ 、つまり、NT が選択される $P(H)$ は $P(H) = .83$ 付近より大きいときとなった。 $P(H) > .50$ でも PT が選択される理由は、PT 時の S_4 の尤度が 0 であることから説明できる。PT を用いた場合、 $P(S_4) = 0$ とすることができるため、NT を用いた場合と比較してエントロピーをより低減できる。その結果として $P(H) > .50$ でも PT を用いることが有効となる。図 4 に c と $P(H)$ を変動させたときの $P(S_0)$ と検証タイプの変遷を示す。

特徴 (3) はフェーズ移行したのちに、多くの検証をくり返さずに検証が打ち切られることを示唆する。解決者が仮説の確信度を一定以上得たときにそれ以降の検証を打ち切ったり、仮説の確信度が急激に上昇することによって十分に検証したとみなしたりすることが考えられる。フェーズ移行に伴う $P(S_0)$ の急激な上昇については、PT 時には S_0 と S_1 が、NT 時には S_0 と S_2 が区別できないことに起因する。フェーズ移行後にこれらが区別できるようになるため、 $P(S_0)$ の急激な上昇と他方の下降が実現す

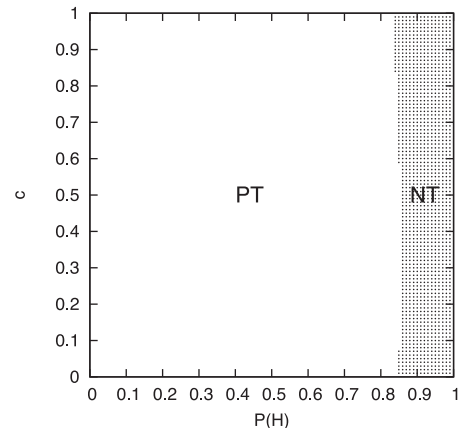


図3 パラメータと初回検証タイプ。ドットは初回検証タイプが NT であったことを示す。

る。このとき、 c が確信度の上昇に影響を与える。 c が大きくなるにつれて $P(FB_P|S_1)$ や $P(FB_P|S_2)$ は大きくなり、確信度の上昇は少なくなる。逆に c が小さくなるにつれ、確信度の上昇は急激となる。一方、 $P(H)$ には依存しない。このことは、図 4 に示した $P(S_0)$ の変遷からもわかる。

特徴 (4) も c に依存し、 $P(H)$ には依存しない。 c は、特徴 (1) の後半にある各フェーズの長さに影響を及ぼす。各フェーズは、 c の値が大きくなるにしたがって長く続き、小さくなると短くなる。その結果、PT と NT の入れ替わりは、 c が小さいほど検証方略の早い段階であられる。換言すると、 c が小さくなるに伴い、現在行った検証タイプの有用性が早く減退する（図 4 の選択される検証タイプの変遷を参照）。

特徴 (5) は、われわれが、どのような（大きさの）仮説についても、ほぼ一定回数で仮説の検証を終えることを示唆している。つまり、解決者がどのような仮説を生成しても、その仮説の正誤を判断するのに必要な検証回数に大きな違いはない。図 4 から、 $P(H)$ の大きさが仮説の確信度 $P(S_0)$ の変遷に大きな影響を与えていないことがわかる。

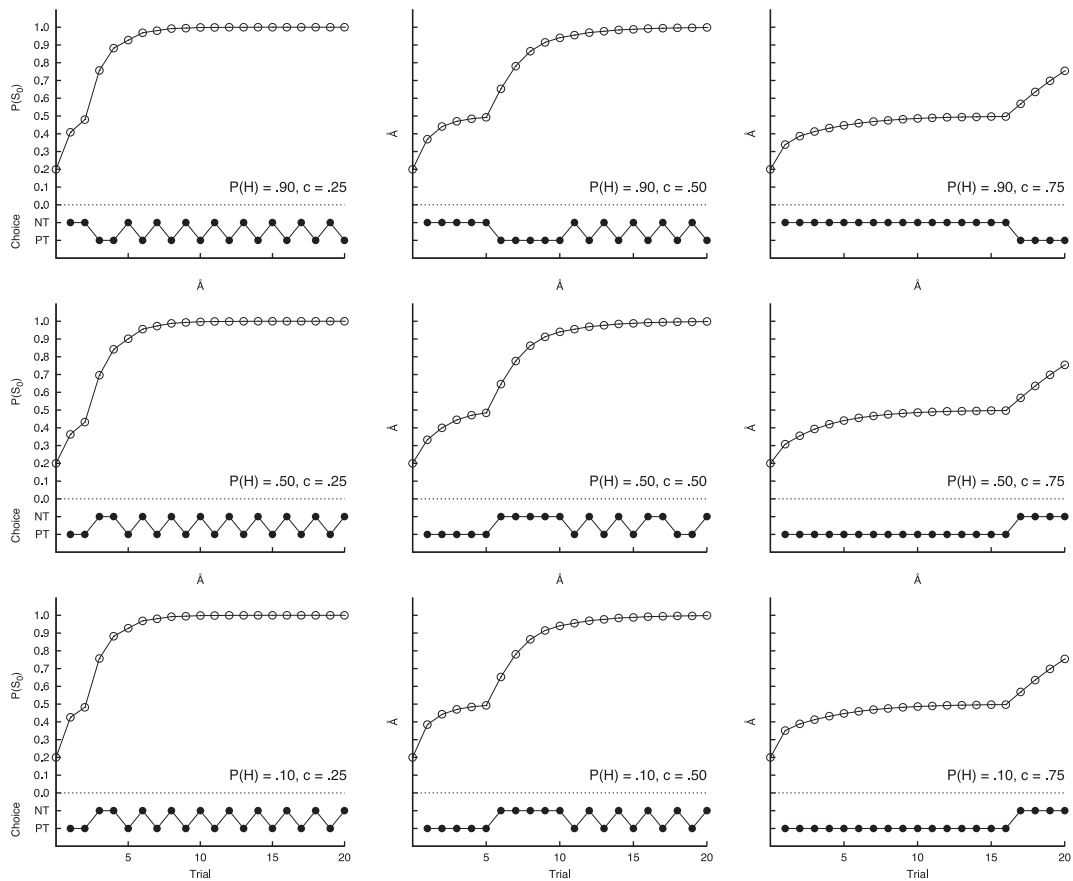


図4 c と $P(H)$ を変動させたときの $P(S_0)$ と選択される検証タイプの変遷。左から右へ向けて $c = .25$, $c = .50$, $c = .75$, 下から上へ向けて $P(H) = .10$, $P(H) = .30$, $P(H) = .90$ と変動させた。

以上の特徴の中で、特徴(2)、(3)はPT方略が適応の合理性を有することを意味する。つまり、多くの局面でPTを用いることが有効であることに加え、PTフェーズから検証を開始した場合、NTフェーズの途中で検証が打ち切られる場合には、結果的に検証系列におけるPTの割合が増えることになる。このことは、PTが多用されることについて、Bayes的規範の観点から合理的な根拠があることを意味する。

3. 実験

シミュレーションの結果から確認された5つの定性的特徴が、実際に観察されるかどうかを調べるために実験を行った。実験はWason (1960) の2-4-6課題を、次の2点で変更して用いた。第1に、オリジナルの2-4-6課題は、問題解決者が規則を発見する規則発見課題であったが、本実験では参加者に

あらかじめ規則を提示し、提示した規則の検証を求める規則検証課題とした。これは参加者に規則を提示することで、検証すべき規則に合う事例が全事例に占める割合(仮説の大きさ)を統制するためである。本研究では、仮説が確認されていく過程を調べることが目的であり、想定される仮説の大きさや仮説を支持する(あるいは支持しない)結果の与え方を統制する必要があるため、このような課題設定とした。第2に、仮説に現れる3つの数字の値は0から99までの正数とした。この変更は、仮説の大きさを厳密に規定するために導入した。

3.1 方法

参加者 大学生および大学院生25名(男4名、女21名)が実験に参加した。平均年齢は22.2歳であった。すべての参加者はボランティアで実験に参加した。

表 3 実験で使用された仮説

仮説 H	仮説の大きさ $P(H)$
稀少事例規則	
すべて偶数	0.13
上昇する数列	0.16
合計が 6 の倍数	0.17
一般事例規則	
積が偶数	0.88
異なる 3 つの数	0.97

注)「仮説の大きさ」とは、本実験で扱った範囲(各数値が 0 から 99 のいずれか)で、実際に仮説を満たす事例の割合を示す。

課題および要因計画 Miwa (2004) で用いられた 3 つの数字に関する規則から、次のような基準で 5 つの規則を抽出した。まず、各規則を満たす実際の事例数を数え、それを全事例数で割ったものを、仮説の大きさ $P(H)$ とした。次に、稀少事例規則 ($P(H) < .20$) と一般事例規則 ($P(H) > .80$) に相当する仮説を合計 5 つ選んだ。表 3 に本実験で用いた仮説(規則)を示した。

稀少事例規則と一般事例規則を実験条件とし(以下では、稀少条件、一般条件)、1 要因 2 水準の参加者内計画とした。

手続き 参加者は課題についての教示を受け、1 セッション 1 問からなる課題を 3 セッション実施した。各セッションは、参加者と実験者が 1 対 1 で対面し、以下の手順で行われた。

はじめに、参加者に検証すべき規則(表 3 の中の 1 つ)を提示した。規則は A5 サイズの用紙の中央に記されており、該当セッション中は常に参加者が確認できるようにした。その後、参加者に具体的な 3 つの数字によって構成される事例を生成し、提示された仮説が正しいかどうかを検証することを求めた。実験者は、参加者が生成した事例が本当に正しい規則に含まれるか否かをフィードバックした。実際には、仮説を支持する結果(PT 時には“YES”, NT 時には“NO”)しか与えられなかった。これは、仮説が確認されていく過程を調べることを目的としたためである。ただし、本セッションで必ず仮説を支持する結果が来ることを参加者に予告させないようにするため、後述のダミー・セッションにおいては、仮説を支持しない結果(PT: “NO”, NT: “YES”)も与えられた。フィードバック後、参加者

にどれくらい仮説が正しいと考えているかを、視覚アナログ尺度で記すことを求めた。視覚アナログ尺度は、左端を仮説が「正しくない」、右端を「正しい」とした 92 mm の水平直線上に、仮説が正しいと思う程度を目分量で印を入れるものであった。検証ごとに、参加者は生成した事例と、視覚アナログ尺度の評定を所定の用紙に記入し、実験者に手渡した。このとき、参加者が仮説が正しい(あるいは誤っている)と十分に確信を持って判断できたら、実験者にその旨を伝えて 1 セッションを終えた。判断できない場合には、この手続きを判断ができるようになるまで続けた。

課題の規則は、稀少条件と一般条件から 1 つずつ無作為に選択し、提示順序はカウンターバランスをとった。これら 2 セッションに先立ち、すべての参加者がダミー・セッションをはじめのセッションとして行った。ダミー・セッションでは、「合計が偶数」($P(H) = .50$)という仮説(規則)が与えられた。ダミー・セッションでは、3 回目以降の検証では仮説を支持しない結果が返され続けた。

参加者あたりの実験時間は、教示を含め、概ね 20 分程度であった。

3.2 モデル適合度の指標

モデルの出力(検証)系列と人の検証系列を比較して、最も適合度が高くなるようにパラメータ $P(H)$, c を決める必要がある。そこで、モデルのデータに対する適合度の指標が必要になるが、既存のものに適当なものはないため、このモデルに合った指標を構成する必要がある。その際に留意すべき点は、検証系列の中に相違があるときにその位置が重要である点と、前の検証結果が後の検証系列に影響を及ぼす点である。以下では、例を使ってこの点を説明する。

検証系列を $\{P, N, P\}$ のように表記することとする。この例は、1 回目 PT, 2 回目 NT, 3 回目 PT という検証系列を表す。たとえば、データ $\{P, P, P\}$ に対して、2 つの検証系列 $\{N, P, P\}$ と $\{P, P, N\}$ は、どちらも 1 箇所の違いであるが、検証系列の最初の方が確信度の変動が大きいことを考えれば、後者の方がデータに近いと考えるべきであろう。

この点を念頭に置いて、適合度の指標を次のように定めることとした。検証系列内における各検証の差異のインパクトを、エントロピー減少量の差

(deviation in entropy reduction; DE) であらわすこととする。すなわち、データがモデルに合わない場合、合わない部分の検証の影響力の分だけモデルにペナルティを課するという考え方である。検証系列の最初から順にこの要領でペナルティを累積する。ただし、DE を計算するたびに、モデルではなくデータが示す検証を実施して状態確信度を更新する。こうしてデータの検証系列の最後まで DE を計算して累積したものを、エントロピー減少の累積偏差 (cumulative deviation in entropy reduction; CDE) と呼ぶ。

より具体的に説明するために、以下に例を挙げる。たとえば、ある参加者の検証系列が $\{P, N, P\}$ であったとする。この検証系列と、パラメータが $P(H) = .50$, $c = .50$ の場合のモデルの検証系列 $\{P, P, P, \dots\}$ を比較する。図 5 に、このモデルから計算されたエントロピーが示されている。参加者が行った 1 回目の検証は PT であり、エントロピー減少量は $2.32 - 1.92 = 0.40$ である。これは、NT を選択した場合 ($2.32 - 2.24 = 0.08$) よりも減少量が多いので、モデルに一致する。1 回目の検証に関しては、モデルとデータの違いがないので、 $DE_1 = 0$ である。しかし、参加者の 2 回目の検証は NT であり、これはモデルと異なる。2 回目の検証では、PT と NT で得られるエントロピー減少量の差は $DE_2 = (1.92 - 1.72) - (1.92 - 1.84) = 0.12$ となる。2 回目の検証は、PT の方がエントロピー減少量が多いが、データは NT なので、NT で状況確信度を更新する。参加者の 3 回目の検証は PT であり、エントロピー減少量は $1.84 - 1.64 = .20$ となる。ただし、NT を選択した場合もエントロピー減少量は (ほぼ) 同じなので、 $DE_3 = 0.00$ となる。したがって、この適合度の指標では 3 回目の検証では PT と NT のどちらを選択しても CDE に影響はない。以上より、 $CDE = \sum_i DE_i = .12$ となる。本研究では、以上のような手順で算出した CDE をモデルのデータとの乖離の程度を示す指標とした。

3.3 結果および考察

5 名の参加者が稀少あるいは一般条件において、仮説を支持する結果を受け取っているにもかかわらず、提示された規則が誤りであると判断した。これらの参加者は課題で要求されていることを勘違いしていたか、提示された規則の代わりに他の規則を検

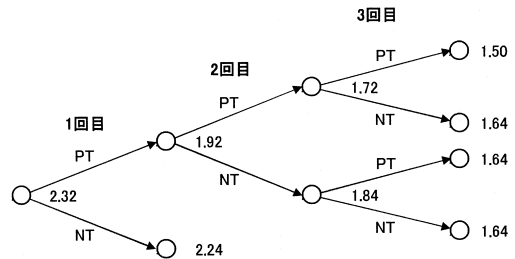


図 5 $P(H) = .50$, $c = .50$ 時の 3 回目の検証までの検証とエントロピーの変遷。数値は該当検証が選択された場合のエントロピーを示す。

証していた可能性があるため、この 5 名を除いた 20 名の実験結果を分析対象とした。

課題の性質 条件ごとに、参加者が選択した検証タイプの平均値を求めたところ、稀少条件では PT は平均 3.20 回 ($SD = 1.33$)、NT は平均 2.10 回 ($SD = 1.45$) であり、一般条件では PT は平均 3.05 回 ($SD = 1.80$)、NT は平均 1.65 回 ($SD = 1.31$) であった。条件と検証タイプで分散分析を行ったところ、検証タイプの主効果のみ有意差が確認された、 $F(1, 19) = 7.09$, $p < .05$ 。条件の主効果、 $F(1, 19) = 2.25$, $p = .15$ 、および交互作用、 $F(1, 19) = .41$, $p = .53$ 、は確認されなかった。これは、参加者が条件に関わらず PT を多用していたことを示している。

モデルの特徴と結果の比較 モデルが示した特徴 (1) については、ほぼ確認された。まず、 $P(H)$ が小さい場合には検証タイプは PT から開始されることを確認するため、初回に選択された検証タイプを分析した (2 項検定)。PT から開始した参加者の割合は、稀少条件では 75% ($15/20$, $p < .05$)、一般条件では 65% ($13/20$, $p = .26$) であった。また、稀少条件では 90% ($18/20$, $p < .01$)、一般条件では 75% ($15/20$, $p < .05$) の参加者が課題実行中に検証タイプの切り替えを行った。

特徴 (2) についても、上でみた開始検証タイプの条件間の違いが、間接的な支持を与えている。すなわち、稀少条件 ($P(H) < .50$) では参加者は検証系列を PT から開始したが、一般条件 ($P(H) > .50$) では有意な差は認められなかった。このことから、初回検証タイプの選択が PT から NT へ切り替わるのが $P(H) > .50$ であること、すなわち、 $P(H)$ の

大きさに対して PT と NT は非対称であること²⁾が示唆される。ただし、モデルでは $P(H) > .83$ 付近では NT から開始されることを予測していたが、一般条件では NT から開始されるという明確な結果は得られなかった。このような結果になった 1 つの理由は、一般条件が PT と NT が選択される境界付近に設定されていたためと考えられる。別の可能性については、総合考察で論じる。

特徴 (5) は間接的に示された。選択された検証タイプの数と条件に関連が認められなかったのは前述のとおりであるが、これに加えて条件間での検証回数にも差が認められなかった。各条件での平均検証回数は、稀少条件では 5.3 回 ($SD = 1.87$)、一般条件では 4.7 回 ($SD = 1.93$) で、有意差は認められなかった、 $t(19) = 1.50, p = .15$ 。これらは、参加者が提示された規則の正否を判断する上で、必要な検証回数や検証タイプが条件に依存していないことを示している。すなわち、 $P(H)$ にかかわらず、限られた検証回数で仮説の確信度 $P(S_0)$ を十分に高めることを示しており、モデルが示した特徴 (5) を支持する結果であるといえる。

一方で特徴 (3) および (4) については十分に分析することができなかった。これらは参加者の検証回数が少なかったことに因る。特徴 (3) については、フェーズ切り替え時の確信度の上昇分を切り替え前の上昇分と比較するためには、最初の検証フェーズが最低でも 2 試行以上必要となる。しかし、フェーズ切り替えが行われた課題のうち、稀少条件では 50% (9/18)、一般条件では 67% (10/15) の参加者が、1 試行の検証だけで検証タイプを切り替えていた。そのため、分析することを断念した³⁾。特徴 (4)

2) 対称とは $P(H) = .50$ を境界とし、 $P(H) < .50$ では PT から、 $P(H) > .50$ では NT から検証が開始されることを指す。これは、 $P(H) < .50$ のときには PT を用いて検証を開始した方が NT を用いて検証を開始するよりも効率的であるという考え (たとえば、Klayman & Ha, 1987) にもとづく ($P(H) > .50$ のときは逆)。しかし、§2.1 で示したとおり、モデルは PT と NT の検証タイプの切り替えを行う境界を $P(H) = .83$ 付近と予測しており、本研究ではこれを PT と NT が非対称であると表現した。

3) 参考までに、分析可能なデータだけを用いて、検証タイプの切り替えが起こる前までの確信度上昇分の平均と、切り替え時の確信度の上昇分について、条件ごとに比較したところ、稀少条件では切り替え前は平均 $.05$ ($SD = .07$)、切り替え後は平均 $-.06$ ($SD = .20$) であり、一般条件では切り替え前平均 $.09$ ($SD = .05$)、切り替え後平均 $.12$ ($SD = .21$) であった。いずれも有意差は認められなかった、 $t(8) = 1.47, p = .18$, $t(4) = .25, p = .81$ 。

の場合については、そもそも検証タイプの切り替えがフェーズ切り替えによるものなのか、それとも仮説の確信度が十分に高まった後に示されると考えられる入れ替わりなのかを明確に区別することができなかった。これらの特徴を検討するためには、たとえば参加者に求める検証回数を固定するような実験手続きとし、より長い検証系列を取得する必要があるだろう。

CDE による分析 稀少条件では PT から開始されるが、一般条件においてどちらの検証タイプから開始されるか明確ではなかった。条件間の違いはたしかに $P(H)$ の大きさの違いによるものであろうが、一般条件の曖昧な結果は、 $P(H)$ の大きさの影響がモデルの予測ほど大きくない可能性を示唆しているとも見られる。そこで、 c に加えて、 $P(H)$ もモデルのパラメータとしてデータから推定することとした。 $P(H)$ と c をそれぞれ $.01 \leq P(H) \leq .99$, $.01 \leq c \leq .99$ の範囲で $.01$ ずつ変化させながら、実験条件別に CDE を算出した。

まず、全体的な傾向を知るために条件ごとに $P(H)$ と c を上記範囲で変化させたときの CDE の平均推移を求めた。分析の結果、 $.50 < P(H) < .60$ 付近で条件とは無関係に CDE が最小となり、 $c = .99$ で同様に条件とは無関係に CDE が最小となった。 $P(H)$ を変化させたときの全ての c の CDE 平均を図 6 に、 c を変化させたときの全ての $P(H)$ の CDE 平均を図 7 に示した。 $P(H)$ と c が条件とは無関係に類似した変遷をとったことは、参加者が条件に依存しない検証方略を選択したことを示している。また、どちらの条件も $.50 < P(H) < .60$ 付近で CDE が最小となったことは、参加者が $P(H)$ を正確に見積もっておらず、条件とは無関係に概ね $.50 < P(H) < .60$ と見積もっていたことを示している。

次に、以上の結果から詳細に各パラメータを決定し、本実験の結果がモデルでどの程度説明できるのか検討した。 c の増加に伴って CDE は減少し、 $c = .99$ で CDE が最小となることから、 $c = .99$ として $P(H)$ が最小になる値を求めた。その結果、条件に関わらず $P(H) = .50$ で CDE が最小となった。このとき、稀少条件では 75 % の参加者 (15/20) が、一般条件では 65 % の参加者 (13/20) が $CDE \approx 0$ ($CDE < .001$) となった。多くの提示規則で、モデルの予測と人の検証活動の間で乖離が小さく抑えら

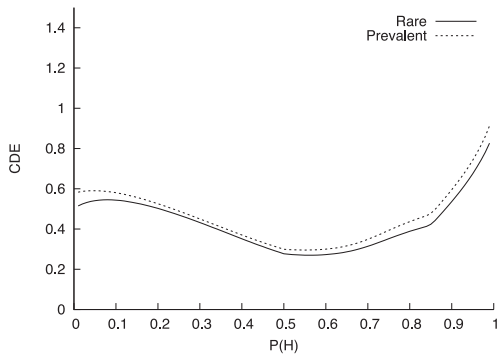


図6 $P(H)$ の変化による条件別 CDE の平均。Rare は稀少条件を, Prevalent は一般条件を指す。

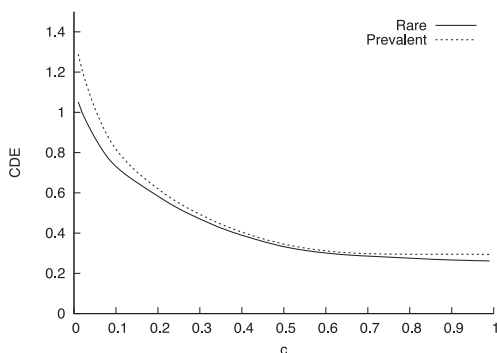


図7 c の変化による条件別 CDE の平均。Rare は稀少条件を, Prevalent は一般条件を指す。

れていることから, 本研究で提案したモデルが人の行動をうまく説明しているといえるだろう。

参加者報告の確信度 視覚アナログ尺度によって報告された確信度は, 左端の「正しくない」を 0, 右端の「正しい」を 1 として, 0 から 1 の範囲を取る値に変換した。初回検証後と最終検証後 (提示された規則が正しいと判断したとき) の間には, 稀少条件, $t(19) = 4.90, p < .01$, 一般条件, $t(19) = 5.59, p < .01$, とともに確信度の上昇がみられた (表 4)。一方, $P(H) = .50, c = .99$ 時のモデルが予測した初回検証後の $P(S_0)$ は .27, 5 回目の検証後は .34 である。これらの結果から, 参加者の自己報告による確信度と, モデルが予測した仮説の確信度が乖離していることが示された。このような結果になった主な原因は, $c = .99$ とすることでモデルでは S_0 と S_1 , あるいは S_0 と S_2 の違いがほとんどなくな

表 4 参加者報告の確信度平均一覧。括弧内は標準偏差。

	初回検証後	最終検証後
稀少条件	.62(.26)	.88(.23)
一般条件	.61(.26)	.94(.10)

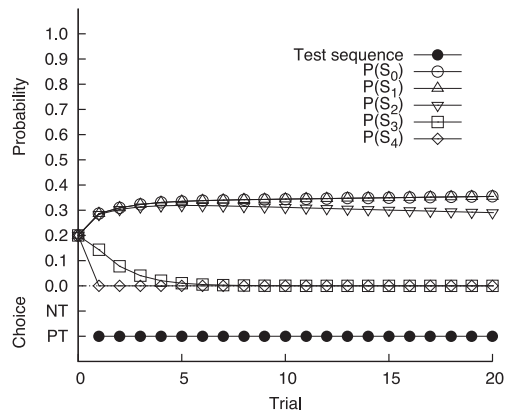


図8 データから推定されたパラメータ ($P(H) = .50, c = .99$) を用いたシミュレーション結果。

ることにあると考えられる。すなわち, S_0 と S_1 , あるいは S_0 と S_2 は類似した確信度を取り, その結果, $P(S_0)$ はあまり上昇しない。これは PT の有用性が下がらないために, その結果, PT をより長い間選択することが原因である。実際, モデルが予測した $P(H) = .50, c = .99$ 時における 15 回目の検証後の $P(S_0)$ は .35 であり, 5 回目の検証後と比較しても, ほぼ横ばいの推移であることが確認された (図 8 参照)。つまり, $P(H) = .50, c = .99$ では仮説の確信度は一定以上上昇しにくいと, 参加者報告の確信度と乖離が生じていると考えられる。

4. モデルの再構築

先の実験では, 参加者は概ね $P(H) = .50, c = .99$ と仮定して仮説を検証していたことが示された。被覆度 $c = .99$ とは, PT においては S_0 と S_2 が, NT においては S_0 と S_1 が限りなく等しくなる状態である。つまり, 実質的に S_0, S_3, S_4 の 3 状況のみを考慮する場合と変わらないことになる。この場合, 被覆度パラメータ c は意味を持たない。したがって, モデルで扱う状況を 3 状況とすることで, より簡潔なモデルを構築することができる。

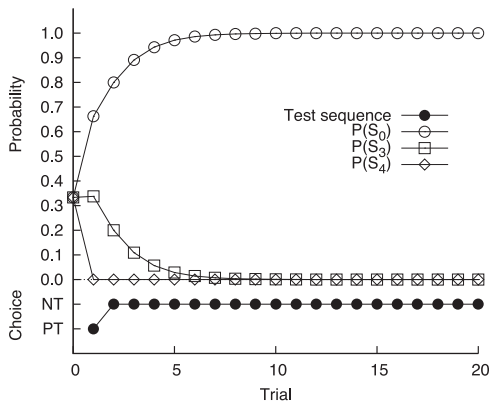


図9 3 状況モデルによるシミュレーション結果 ($P(H) = .51$ の場合)。

4.1 再構築とモデル出力の定性的特徴

3 状況への再構築は単に S_1 と S_2 を除くことで実現した。また、これに伴って各状況の確信度の初期値を $P(S_n) = .33$ とした。3 状況での $P(H) = .51$ のシミュレーション結果を図9に示した。§2.1で挙げた定性的特徴のうち、再構築後に確認された特徴は(1)、(2)、および(5)の3点であった。先の実験で確認されなかった特徴(3)、(4)は、このモデルには見られないことがわかった。

3 状況モデルでも、初回検証が PT から NT へ切り替わるのは $P(H) = .85$ 付近であり、5 状況の場合と変わらない。しかし、3 状況モデルは、5 状況モデルに比べてその振る舞いが明らかに異なっている。まず、 $P(H) \leq .50$ の場合、PT から NT フェーズへの移行は起こらず、NT フェーズから PT フェーズへの移行はいかなる場合も認められない。 $P(H) = .85$ 付近以下の場合、モデルは初回検証では PT を選択する。これは、5 状況のモデルのときと同様に、PT を選択することで $P(S_4) = 0$ とすることができるためである。しかし、2 回目以降の検証では、検証タイプは $P(H) = .50$ を境界として、PT を選択するか ($P(H) < .50$ の場合)、NT を選択するか ($P(H) > .50$ の場合) に固定される。2 回目以降の検証では、モデルの検証タイプの選択に影響を与える確信度は $P(S_0)$ と $P(S_3)$ のみとなるが、 S_0 では PT と NT の尤度が等しいため、実際には S_3 の尤度のみが影響を与える。この結果、常に同じ検証タイプを行った場合の方が $P(S_0)$ が大きく、 $P(S_3)$ が小さくなり、選択される検証タイプの切り替えは起こらない。したがって、特徴(4)

が認められなくなった。なお、 $P(H) = .50$ のときは2回目以降の検証では、PT を行った場合と NT を行った場合のエントロピーは等しくなる。このとき、モデルの仮定により PT が選択される。

また、仮説の確信度の上昇は概して急激であり、少ない検証回数で仮説の正しさを確信する傾向がある。これと関連して、特徴(3)のフェーズ移行にともなう確信度の急激な上昇も認められなくなった。これらはモデルで扱う状況が減ったことが原因である。そのため、1 回の検証が確信度の更新に与える影響が大きくなった。さらに、5 状況モデルでみられた S_0 と S_1 、あるいは S_0 と S_2 を区別する必要がなくなったことが、特徴(3)の消失としてあらわれた。

4.2 実験結果の再分析

先の実験について、CDE による分析と参加者報告による確信度について再分析した。 $P(H) = .50$ のときの CDE は、5 状況 ($P(H) = .50$, $c = .99$) と同様に稀少条件では 75% (15/20) の参加者が、一般条件では 65% (13/20) の参加者が $CDE < .001$ となった。このことは、本実験の参加者たちが用いた検証方略が、より単純化したモデルでも十分に説明可能であることを示している。また、3 状況では5 状況と比較して全体的に CDE は上昇した。これは、モデルを単純化したことによって、1 回の検証が与える影響が大きくなったためである。

3 状況で再構築したモデルでは、初回検証後は $P(S_0) = .67$ 、5 回目の検証後は $P(S_0) = .97$ である ($P(H) = .50$ のとき)。これは、5 状況で構築したモデルに比べるとデータへの適合度がかなり高い。参加者報告平均確信度とシミュレーション結果の比較を図10に示した。条件ごとに、参加者がフィードバック後に報告した確信度と、 $P(H) = .50$ のシミュレーション結果を比較した。両条件で、参加者は初回検証後のフィードバックで、仮説の確信度を $P(S_0) > .60$ と見積もっており、その後、比較的高い確信度の推移を示した。本研究で行った実験手続きでは、検証回数は参加者によって決定されるため、検証を重ねるごとに得られる参加者報告確信度数は減少する。そのため、各検証でのデータの精度にも偏りがあり、正確さは欠くものの、3 状況で再構築したモデルが参加者の確信度をある程度予測しているといえるだろう。また、モデルが予測

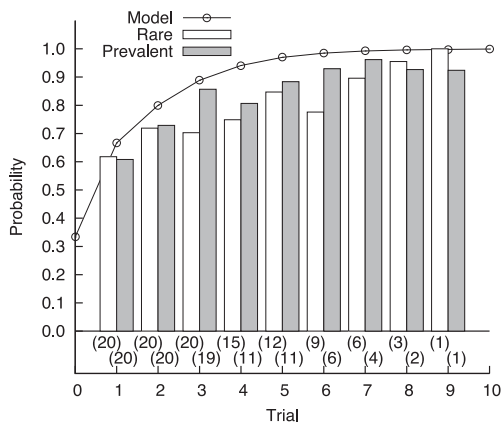


図 10 参加者報告平均確信度とシミュレーション ($P(H) = .50$) 結果. Rare は稀少条件, Prevalent は一般条件の参加者報告平均確信度を, Model はシミュレーションによる仮説の確信度 ($P(S_0)$) を指す. 括弧内は確信度報告が得られた参加者数 (上段: 稀少条件, 下段: 一般条件).

した仮説の確信度を母平均と仮定して, 参加者報告による確信度と母平均の差の検定を行ったところ, 稀少条件初回検証後, $t(19) = 0.81, p = .43$, 稀少条件最終検証後, $t(19) = 1.67, p = .11$, 一般条件初回検証後, $t(19) = 1.00, p = .33$, 一般条件最終検証後, $t(19) = 1.22, p = .24$ のすべてにおいて有意差は認められなかった. この結果はモデルの予測が参加者の行動と矛盾しないことを示している.

前述のとおり, 3 状況で再構築したモデルでシミュレートした場合, 5 状況のモデルと比較して確信度の上昇は急激である. そのため, 一定以上の確信度を得るまでの必要な検証回数は減少する. 本実験でも, 参加者たちの平均検証回数は稀少条件で 5.3 回, 一般条件で 4.7 回であり, 検証回数が多いとはいえない. この結果も, 3 状況モデルを支持している.

以上より, 実験参加者たちは, 5 状況よりも 3 状況を想定して提示規則を検証したと考えるのが妥当であろう. しかし, S_1 や S_2 を考慮しないのは, 仮説検証としては十分なものではない. すなわち, 3 状況モデルは, 5 状況モデルと比較して, より単純な, ヒューリスティックに基づく仮説検証のモデルであるといえる.

他の 2-4-6 課題の実験と比較して, 本実験の参加者たちが単純な仮説検証を行う傾向が強かったのかどうかはわからないが, もしその可能性があるとする

れば, 仮説自体が実験者から与えられたものであることが関係しているかもしれない. すなわち, 仮説が自分で生成したものではなかったことの影響があらわれている可能性がある. たとえば, 提示された仮説では, 自分で生成した仮説に比べて思い入れが少ないなどの理由で, 仮説検証の手続きに多くの労力を費やして詳細に調べようとする動機づけが低いということがないとは言いきれない. この点については本研究の範囲を超えるため, これ以上の考察は控えるが, 今後の研究において留意すべき点の 1 つかもしれない.

5. 総合考察

本研究では, まず, 期待獲得情報量最大化の原理と Bayes 的確信度更新に基づく仮説検証過程のモデルを構築した. 次に, 2-4-6 課題 (Wason, 1960) の変形版を用いた実験を行い, そこで得た参加者のデータとモデルの振る舞いを比較したところ, モデルの特徴の一部がデータに認められなかった. そこで, データからモデルのパラメータ推定を行ったところ, 課題中で与えた仮説の大きさ $P(H)$ に関わりなく, モデルは $P(H) = .50$ 付近で最も高いフィットを示した. また, H と T の大きさの関係をあらわすパラメータ c は, ほぼ 1 と推定された. その結果を踏まえてモデルを再検討したところ, 参加者の仮説検証過程は, かなり単純なモデル (3 状況版) でもよく適合することが明らかになった. 以上の結果は, 人々の仮説検証過程が, 概ね Bayes 的規範に従うものであることを示すと同時に, 完全な規範 (5 状況版) ではなくヒューリスティック (3 状況版) に基づく過程であることを示しているといえるだろう.

CDE に基づくモデルフィットでは, いずれの条件のデータに対しても, 課題中で与えた仮説の大きさ $P(H)$ に関わりなく, モデルは $P(H) = .50$ 付近で最も高いフィットを示した. このことは, 参加者が, 与えられた仮説の大きさを「正しく」見積もっていなかったことを示す. 「正しく」というのは, 事例を 2 桁の正数のみに限定して近似的に定義した $P(H)$ に合致することである. 実際には, 負の数や 3 桁以上の数も仮説の対象となり得ることから, 事例の定義域を有限の集合に限定することはできない. したがって, 定義域の曖昧さが, このような結果の一因になっていると考えることも可能である.

もう一つ考えられる要因は、仮説の大きさの情報に敏感になりすぎるものの費用対効果である。もし検証によって得られる効用にほとんど差がないとすれば、計算を正確にするためにコストをかけても意味がない。つまり、 $P(H) = .50$ がデフォルト値として半固定化していることが、別のレベルでの最適化になっている可能性がある。実は、3 状況モデルでは、確信度の上昇が急激であることから容易に推定されるように、初めの数回以降の検証は、PT でも NT でもエントロピー減少量はほとんど変わらない。特に、 $P(H) = .50$ のときは、初回の検証は PT だが、2 回目以降は、PT でも NT でもエントロピー減少量は全く同じである。その上、モデルの定性的特徴 (5) は 3 状況モデルでも見られ、このような事態が $P(H)$ の値によって大きく変化しないことを示している。

実際には、計算をより正確にするためにかかる認知的コスト自体を正確に見積もるのは困難である。また、計算をより正確にすることによって得られる期待獲得情報量は容易に計算できるが、検証される仮説が現実世界においてもたらず意味を正確に見積もるのはさらに困難である。したがって、与えられる情報に対してどれくらい鈍感であるのが最適であるのか、という点を正確に推定することは極めて難しい。ただ、人々が与えられた情報に基づく信念更新において、Bayes 的規範に比べて保守的であることは古くから知られてきた (Edwards, 1968) が、この知見は、仮説の大きさの情報に敏感になりすぎないことと無関係ではないように思われる。つまり、 $P(H) = .50$ をデフォルト値として半固定化することが、ある種のヒューリスティックとして機能している可能性がある。この点については、本研究のみから結論することは困難であり、今後、様々な課題を用いて証拠を集め、検証していく必要があるだろう。

ただ、以上の議論は、外部情報によって $P(H)$ のデフォルト値を上書きすることの困難さに関するものであり、特定のデフォルト値の有効性を含意するものではない。 $P(H) = .50$ であることは、日常生活における効率的な仮説の取捨選択方略と関係があるかもしれない。全事例を同数ずつに 2 分割 ($P(H) = .50$) するような仮説は、最もエントロピーが高く、仮説を限定しながら外延を絞っていく際に立てる仮説としては最も効率的である。たとえ

ば、相手が考えている 100 以下の正数を当てる場合、「50 以下の数ですか」というように、事例数が 2 分割されるように質問を続けていくのがよい。本実験から推定された $P(H)$ は、このような効率的な仮説の取捨選択方略が反映されたものかもしれない。

実験の結果、データは、3 状況しか想定しない単純なモデルで十分に説明されることがわかった。この結果は、他の推論課題や確率判断課題で広く認められている等確率の仮定に関する仮説 (服部, 2008) と符合して興味深い。この仮説によれば、われわれは、考慮対象となっている 2 事象のデフォルト確率がほぼ等しいと仮定する傾向を持つ。ターゲット集合 (T) が仮説集合 (H) と同サイズであれば、理論的に 3 状況しかないことになる。3 状況モデルで十分にデータの記述ができたことは、参加者が等確率を仮定していたことを示唆する。また、多くの参加者が 2-4-6 課題の初期場面で戸惑うことも、3 状況モデルの記述的妥当性を裏付けている。3 状況モデルは、ヒューリスティックとして等確率を仮定している状態を端的に表しているといえることが可能である。

しかし、等確率の仮定をどこかで崩さない限り、2-4-6 課題の正解にたどり着くことはできない。実際には、多くの参加者が最終的には正解に至ることから、現実的な仮説検証過程を扱うためには 5 状況モデルが必要であることは明らかである。Gorman & Gorman (1984) は、規則発見課題における正解者は、PT を用いて検証を開始し、後に NT を併用した検証方略を用いる傾向があることを指摘した。仮説検証の初期では PT を用いて仮説の蓋然性を高め、終盤では NT を併用して仮説を洗練すると考えられている。このような現象は、まさに 5 状況モデルで確認された特徴 (4) である。つまり、3 状況モデルはヒューリスティックな仮説検証過程を、5 状況モデルはより詳細な検証過程を表現しているとみなすことができる。

以上を踏まえると、人々の仮説検証過程には、Bayes の定理に基づく連続的な信念更新の過程とは別に、モデル間の移行のような非連続的な過程が存在すると考えるべきであろう。本研究で提案した 5 状況版のモデルは、 c や $P(S_n)$ の初期値を操作することで、モデル間の移行を連続的に捉えることが可能⁴⁾であり、技術的には、5 状況モデルの枠組みで 3 状況モデルを扱うことは困難ではない。し

かし、ヒューリスティックな過程から分析的な過程への移行まで含めて、仮説検証過程の全体像を単一の確率的モデルで整合的に表現できるかどうかは、現時点では明らかではない。

本研究では、複数の仮説を同時に扱う診断検証(DT)は扱わなかったが、今後は、モデルに取り入れていく必要があるだろう。これまでの研究で、現実の仮説検証過程において、DTは比較的良好に用いられることが知られている。Klayman & Ha (1989)は仮説検証過程の中盤まで、DT(彼らの表現では alternative test)が用いられることを示した。また、Tukey (1986)も仮説検証過程の初期では、仮説を取捨選択するための検証が行われることを指摘した。われわれは複数の仮説を保持し、仮説を効率的に検証している。したがって、より人間の仮説検証過程に近いモデルを構築するためには、DTを無視することはできないだろう。さらには、本研究のモデルでは仮説自体の生成過程は扱わなかったが、仮説生成という創造的活動も大きな課題である。今後は、仮説検証過程全体を表現するモデルが期待されよう。

6. 結 論

本研究では、仮説が支持され続ける場合の検証系列を扱う比較的単純な Bayes モデルを構築した。このような単純なモデルにもかかわらず、モデルは直観的にも理解しやすいわれわれの検証活動を予測、説明した。実験では、解決者が実際の仮説に依存せずに仮説の大きさを $P(H) = .50$ と見積もっていることが示された。さらに、われわれが仮説検証を行う際に、本研究の参加者が示したように、ヒューリスティックな検証方略を用いることがあることが示された。

しかし、ヒューリスティックな仮説検証と分析的な仮説検証がどのように使い分けられているのかなど、解明していくべき疑問点も残っている。これらは、今後明らかにしていくべき課題である。

謝 辞

本研究の実施に際して、眞嶋良全氏から大変有用なコメントを頂いた。ここに記して感謝する。

4) 本稿では議論しなかったが、5 状況版から 3 状況版へ再構築する際に、被覆度 $c \rightarrow 1$ とする以外に、初期信念確率 $P(S_1), P(S_2) \rightarrow 0$ とすることで、3 状況版へ移行することができる。

なお、本研究の実施と本稿の執筆は、日本学術振興会科学研究費補助金 19500229 (研究代表者: 服部雅史) の資金援助を受けた。

文 献

- Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: John Wiley & Sons.
- Fine, T. L. (1973). *Theories of probability: An examination of foundations*. New York: Academic Press.
- Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review*, **100**, 298–319.
- Gorman, M. & Gorman, M. (1984). A comparison of disconfirmatory, confirmatory and control strategies on Wason's 2-4-6 task. *The Quarterly Journal of Experimental Psychology*, **36A**, 629–648.
- 服部雅史 (2008). 推論と判断の等確率性仮説: 思考の対称性とその適応的意味. 『認知科学』, **15**, 408–427.
- Hoenkamp, E. (1989). 'Confirmation bias' in rule discovery and the principle of maximum entropy. *Proceedings of the 11th Annual Conference of the Cognitive Science Society*, 651–658.
- Klayman, J. & Ha, Y.-W. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, **94**, 211–228.
- Klayman, J. & Ha, Y.-W. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 596–604.
- Miwa, K. (2004). Collaborative discovery in a simple reasoning task. *Cognitive Systems Research*, **5**, 41–62.
- Nisbett, R. & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Oaksford, M. & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, **101**, 608–631.

- Tukey, D. D. (1986). A philosophical and empirical analysis of subject's modes of inquiry in Wason's 2-4-6 task. *The Quarterly Journal of Experimental Psychology*, **38A**, 5-33.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Psychology*, **12**, 129-140.
- Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology*, 135-151. Harmondsworth, Middlesex, UK: Penguin Books.

(Received 8 July 2009)

(Accepted 8 Dec. 2009)



田村 昌彦 (学生会員)

2007 年立命館大学文学部卒業。2009 年名古屋大学大学院情報科学研究科修士課程修了。現在、名古屋大学大学院情報科学研究科博士後期課程に在籍。推論、意思決定、洞察問題解決などの高次認知機能に興味がある。日本認知科学会、日本心理学会、各会員。



服部 雅史 (正会員)

1990 年北海道大学文学部卒業。1996 年同大学大学院文学研究科博士後期課程単位取得退学。博士(文学)。1997 年より立命館大学文学部。2003-2004 年英国(ウェールズ)カーディフ大学心理学部客員研究員。現在、立命館大学文学部教授。推論、意思決定、問題解決などの高次認知機能の研究に従事。人間の思考・言語・コミュニケーションの適応的合理性に興味を持つ。日本心理学会(2007- 編集委員)、日本基礎心理学会、Cognitive Science Society, Society for Judgment and Decision Making ほか会員。hat@lt.ritsumei.ac.jp



三輪 和久 (正会員)

1984 年名古屋大学工学部卒業。1989 年同大学大学院工学研究科博士課程修了(情報工学専攻)。工学博士。1989 年同大学情報処理教育センター助手、1993 年同大学大学院人間情報学研究科助教授を経て、2004 年より名古屋大学大学院情報科学研究科メディア科学専攻教授。1991 年から1992 年、米国 Carnegie Mellon University, Dept. of Psychology, visiting assistant professor。認知科学、人工知能、教育工学の研究に従事。とりわけ、発見、創造、洞察、協同など、人間の高度思考過程に興味がある。

付 録

A. 尤度の算出

各状況の尤度 $P(FB_P|S_i) = P(T|H, S_i)$ は以下のように算出した。

A.1 PT の場合

状況 S_0 および S_1 では、常に仮説を支持する結果が返るため、

$$P(FB_P|S_0) = 1, \quad (4)$$

$$P(FB_P|S_1) = 1. \quad (5)$$

状況 S_2 では、 $P(H, T) = P(T) = cP(H)$ ゆえ、

$$\begin{aligned} P(FB_P|S_2) &= P(T|H, S_2) = \frac{P(H, T|S_2)}{P(H|S_2)} \\ &= c. \end{aligned} \quad (6)$$

状況 S_3 では、 H, T は独立で $P(H) = P(T)$ ゆえ、

$$P(H, T) = P(H)P(T) = P(H)^2.$$

よって、

$$P(FB_P|S_3) = \frac{P(H, T|S_3)}{P(H|S_3)} = P(H). \quad (7)$$

状況 S_4 では、常に仮説を支持しない結果が返るので、

$$P(FB_P|S_4) = 0. \quad (8)$$

A.2 NT の場合

状況 S_0 および S_2 では、常に仮説を支持する結果が返るため、

$$P(FB_P|S_0) = 1, \quad (9)$$

$$P(FB_P|S_2) = 1. \quad (10)$$

状況 S_1 では、 $P(\bar{H}, \bar{T}) = P(\bar{T}) = c$ ゆえ、

$$\begin{aligned} P(FB_P|S_1) &= P(\bar{T}|\bar{H}, S_1) = \frac{P(\bar{T}|S_1)}{P(\bar{H}|S_1)} \\ &= c. \end{aligned} \quad (11)$$

状況 S_3 では、 H, T は独立で $P(\bar{H}) = P(\bar{T})$ ゆえ、

$$P(\bar{H}, \bar{T}) = P(\bar{H})P(\bar{T}) = P(\bar{H})^2.$$

よって、

$$\begin{aligned} P(FB_P|S_3) &= \frac{P(\bar{H}, \bar{T}|S_3)}{P(\bar{H}|S_3)} = P(\bar{H}|S_3) \\ &= 1 - P(H). \end{aligned} \quad (12)$$

状況 S_4 では、 $P(\bar{H}, \bar{T}) = 1 - P(H) - P(T)$ 。また、 $P(T) = P(H)(1 - P(H))$ と定めたため、

$$\begin{aligned} P(\bar{H}, \bar{T}) &= 1 - P(H) - P(H)(1 - P(H)) \\ &= (1 - P(H))^2. \end{aligned}$$

よって、

$$P(FB_P|S_4) = \frac{P(\bar{H}, \bar{T}|S_4)}{P(\bar{H}|S_4)} = 1 - P(H). \quad (13)$$