仮説検証における事例選択の変化^{1), 2)}

眞嶋良全 (北海道大学)

Selection of test cases within hypothesis testing

Yoshimasa MAJIMA (*Hokkaido University*)

The present study investigates whether the selection of test cases changes during hypothesis testing. In total, 75 participants solved one of two discovery tasks: the 2-4-6 task or the New Eleusis task. The participants were first presented with an initial instance produced by a particular rule and were asked to determine the rule. On each trial attempt, after the participants had formulated a hypothesis about the rule and proposed a new instance, they were provided with feedback about whether or not the instance fit the rule. The task would end when the participants announced that they were absolute confident in their best guess. The results indicate that the selection of alternative tests decreased in the latter phases, although the selection of negative and positive tests increased with time. In particular, successful problem solvers were much more likely to produce negative tests. These findings suggest that the process of hypothesis testing consists of at least two different phases.

Key words: hypothesis testing, rule discovery task, selection of test cases

本研究は、仮説検証過程における検証事例の選択が、検証過程を通じて変化するかどうかを検討するために行われた。75名の実験参加者は、2-4-6問題、および新エリュシス課題とよばれる規則発見課題のいずれかを解いた。参加者は、初期事例を与えられた後で、その事例を含む規則を発見するよう求められた。参加者は、毎試行において、その時点で考えられる仮説を挙げ、検証事例を選ぶよう求められた。また、検証事例が規則に合致するかどうかフィードバックされ、参加者は規則を発見したと確信が持てたときに、それを報告するよう求められた。その結果、仮説検証の終盤において、複数の仮説に対する相互排他的な検証を行う診断テストの選択が減少する一方で、仮説に沿った正事例検証、および仮説に沿わない負事例検証の選択は増加することが示された。負事例検証の選択の増加は、特に正解到達者において顕著であった。この結果から、検証の選択は、仮説検証過程を通じて固定したものではなく、少なくとも序盤と終盤では、異なった方略に基づいて選択が行われていることが示唆される。

キーワード: 仮説検証, 規則発見課題, 検証選択

人は日常生活の中で、限られた経験の中からさまざまな規則を発見し、物事の説明や他者の行動の予測を行う。このような認知的活動が帰納推理である。帰納推理は個々の具体的な事象から一般的規則を導く推論形式であるが、心理学的過程としては、事例の観察、仮説の生成、検証事例の選択、結果の評価という段階が区別され、これら一連のプロセスを仮説検証過程と呼んでいる。仮説検証過程分析の重要性は、科学的探求状況のみならず、社会的認知をはじめとするさまざまな領域において指摘されている。

仮説検証過程に関する心理学的研究では、人が、仮説

に合致した、あるいは仮説を支持する事例による検証を数多く行うことが示されてきた (Klayman & Ha, 1987, 1989; Mynatt, Doherty, & Tweney, 1977, 1978; Snyder & Swann, 1978; Tweney, Doherty, Warner, Pliske, Mynatt, Gross, & Arkkelin, 1980; Wason, 1960)。例えば、代表的な実験課題である2-4-6問題 (Wason, 1960) では、実験参加者は、“2, 4, 6”という数字列を提示される。参加者は、この数字列を決定する規則を発見するよう指示され、その際、規則の候補となる仮説を述べ、その真偽を確かめるために、別の数字列を検証するよう求められる。このとき、“2, 4, 6”という初期事例からは、“連続した偶数”といった初期仮説が生成されやすく、この仮説を検証するために、“10, 12, 14”といった数字列が選択されやすい。ここで、“10, 12, 14”は、“連続した偶数”仮説に合致する正事例であり、この課題を解いた参加者の多くが、正事例検証を多用することが知られている。

この課題において実験参加者が発見すべき規則は、

1) 本研究は、平成17～19年度文部科学省科学研究費補助金・若手研究(B) (課題番号17730422代表: 眞嶋良全) の補助を受けた。

2) 実験実施にあたって、本下史記氏に協力していただいた。また、本論文の草稿段階において、瀧川哲夫先生 (北海道大学) より貴重な助言をいただいた。ここに記して感謝いたします。

“上昇系列” というものであるが、初期仮説である“連続した偶数”は、発見すべき規則の部分集合となっている。仮説が発見すべき規則の部分集合であるときは、仮説の正事例のみを検証しても、仮説を支持する結果しか得られないために、誤った（より厳しい）規則を報告してしまうことになる。一方、仮説に合致しない事例を用いた負事例検証、例えば“連続した偶数”仮説に対する“1, 3, 5”という数字列は、仮説の誤りを示すことができるため、非常に有益な情報をもたらす。このことから、Wasonをはじめとする初期の研究者達は、正事例検証を多用する傾向を、科学的推論の規範理論（反証主義、Popper, 1959）から逸脱した行為ととらえ、確証バイアス (confirmation bias) と呼んだ。しかし、近年の研究では、正事例検証は必ずしも非合理的な方略ではないことが指摘されており、有限な認知資源という制約の下で、獲得できる情報量を最大化することを目的としたヒューリスティックの一種として位置づけられている（例えば、Klayman & Ha, 1987, 1989）。

また、複数の仮説を同時に保持することが課題解決率の向上に貢献することが知られている (Gorman, Stafford, & Gorman, 1987; Tweney et al., 1980; Wharton, Cheng, & Wickens, 1993)。特に、対立仮説を立てた上で、主仮説に対しては正事例、対立仮説に対しては負事例となる検証は、強力推論 (strong inference) と呼ばれる合理的な推論方略であり、情報の持つ診断的価値を重視していることから、診断テスト方略とも呼ばれている (Freedman, 1992; Klayman & Ha, 1989; 三輪, 1999; Platt, 1964)。この点について、Klayman & Ha (1989) は、全体としては正事例検証が最も多く選択されるものの、最終的に規則を発見できた参加者のほうが、発見できなかった参加者よりも、診断テストの選択回数が多いことを示している。

規範からの逸脱ととらえるか、より適応的なヒューリスティックの現れとして考えるかの違いはあるが、人が仮説検証過程において、正事例検証を多く選択することは、この研究領域で繰り返し主張されてきたことである。しかし、ここで、検証の選択は、検証過程を通じて変化のない、固定的なものとして考えてよいのかという疑問が生じる。例えば、初期の段階では、所与のデータから考えられる複数の仮説群から、もっともらしい仮説を抽出することを目的とした診断テストを実行し、いったん仮説が抽出された後では、仮説の蓋然性を高めるための正事例検証や、仮説の不備を検討するための負事例検証を用いる、という方略は直感的にも納得しやすい。もし、われわれが、このような方略に従って仮説検証を行っているのであれば、検証の序盤と終盤では、それぞれの検証の選択率が異なることが予想される。この点について、以下に挙げる2つの研究が、事例選択パターンは固定

されていないことを示している。

Tukey (1986) は、2-4-6 問題の開始直後、および規則報告が行われる直前の試行における検証の選択を比較し、規則報告が行われる直前で、正事例検証が多く行われることを示している。また、Klayman & Ha (1989) は、課題解決の上限として設定した18試行を6試行ごとの3ブロックに分割し、検証選択の変化を検討した。その結果、診断テストは、第1ブロックに比べて第2ブロックで選択がやや増加し、第3ブロックにおいて減少すること、正事例検証はこれとは逆のパターンを示すなど、ブロックによって選択傾向が異なることが明らかとなった。

これらの知見から、人の仮説検証は、検証過程を通じて変化する動的な過程であることが示唆される。しかし、先行研究では、検証の種類の分類基準や、検証過程の段階の設定に問題があることが指摘できる。例えば、Tukey (1986) の検証の分類は、参加者がどのような意図をもって数字列を選択したかの自己報告に基づいており、実際に行われた検証が、報告された意図を実現しているという保証はない。特に、診断テストに相当する“棄却” (Eliminate) を理由として挙げた場合に、実際には主仮説、対立仮説の双方に対して同時に正事例検証を行ってしまい、仮説の棄却に失敗することはありうる。したがって、参加者が選択した検証は、その時点で保持されている仮説との関係によって分類することが適切であると考えられる。

Klayman & Ha (1989) では、棄却に失敗した場合を非診断テストと分類しており、検証の分類基準は改善されている。しかし、試行数を18に固定するという操作が、逆に仮説検証方略の正確な記述の障害となっている可能性がある。例えば、18試行よりも早い時点で、規則を発見したと確信した参加者にとっては、すでに十分な証拠を得たと考える仮説に対して、追加の検証を行う必要はないと考えるかもしれない。そのような参加者は、規則に確信を持った後の試行でも選択を強制されることになるが、それ以前と同じ方略に従って検証を選択するかどうかはわからない。一方で、18試行では正解に至らなかった参加者については、途中で強制的に実験が終了させられてしまうため、18試行を用いた分析が、その参加者の仮説検証過程を完全に記述しているとはいえない。

本研究では、これらの問題を解決するため、参加者が規則を発見できたと確信し、規則を報告するまでを1つの区切りとし、その中での検証選択傾向の変化を、生成された仮説と検証との関係に基づいて分析する。ここで、参加者によって報告に至るまでの試行数が異なることが予想されるため、参加者ごとに課題開始直後からの5試行と、規則が報告される直前の5試行において、後

述する4種の検証の選択頻度を算出し、課題の序盤と終盤における選択傾向を比較する。Klayman & Ha (1989)に従うならば、正事例検証は終盤において増加することが予想され、逆に診断テストは減少することが予想される。負事例検証については、先行研究では、全体的な選択率が低いため分析の対象となっていなかったり、診断テストとひとまとめにされているため、明確な予想はできないが、検証の終盤で仮説の不備を探索するための検証が行われるのであれば、終盤において負事例検証が増加すると推測される。

また、先行研究では、正解到達者のほうが、未到達者よりも診断テストや負事例検証を多く選択することも明らかにされている(Klayman & Ha, 1989; Wason, 1960; Vartanian, Martindale, & Kwiatkowski, 2003)。本研究においても、これらの検証の選択と解決成績との間の関係を明らかにするため、課題の序盤、終盤それぞれにおける4種の検証の選択頻度が、規則発見の成否に与える影響を検討する。

方 法

課題

本実験では、2つの規則発見課題を用意した。1つは、この研究領域の代表的な実験課題である2-4-6問題(Wason, 1960)であり、もう1つは、トランプカードの配列方法に関する規則を発見する新エリュース課題(New Eleusis)である(例えば、Gardner, 1977)。いずれの課題も、コンピュータ上の課題として実施された。

2-4-6問題は、3つの数字から構成される数字列を構成するための規則を発見する課題である。本研究で用いた2-4-6問題で参加者が発見すべき規則は、Wason (1960)と同様「上昇系列」であり、「2, 4, 6」が規則に合致する初期事例として与えられた。参加者は、それぞれの試行で、「10, 12, 14」などの他の数字列を検証し、その検証結果を元に規則を発見するよう求められた。

新エリュースは、トランプカードの配列法に関する規則を発見する一種のゲームであり、科学的発見のプロセスをシミュレートした帰納推理課題である。このゲームでは、ディーラー(本実験では実験者に相当する)が、カードの配列法に関する規則を決め、プレイヤー(本実験では参加者に相当する)は、手持ちの札からカードを出す。このとき、出されたカードが、規則に沿っているものであれば、水平方向に並べられるが、規則に沿わない場合は、垂直方向に並べられる(Figure 1)。本実験では、このゲームを元にした、規則発見課題を作成した(以後、この課題を単にエリュース課題と呼ぶ)。新エリュースは、通常4名程度のプレイヤーで規則発見の早さを競うゲームであるが、本実験で用いた課題は、プレイヤーが実験参加者1名のみである点が異なっ

Figure 1. Example of New Eleusis task.

いた。また、新エリュースでは、プレイヤーはディーラーによって配られた札の中からカードを選択するが、本実験のエリュース課題では、各スート13枚ずつ、合計52枚のカードをすべて手持ちの札として利用することができた。エリュース課題で参加者が発見すべき規則は、「数字が書かれたカード(2~10までのカード)」というものであり、ハートの3、スペードの5、ダイヤの7が規則に合致する初期事例として与えられた。参加者は、それぞれの試行で、52枚のカードの中から任意の1枚を検証し、その検証結果を元に規則を発見するよう求められた。

手続き

実験参加者は、課題の目的、各試行における手続きに関する教示を与えられた後、PCの操作に習熟するための練習試行を実施した。練習試行では、2-4-6問題、エリュース課題ともに、どのような検証を行っても、常に規則に合致するというフィードバックが与えられた。実験開始後は、参加者は初期事例を与えられ、以後の各試行で、考えられる規則の候補、すなわち仮説を1つ以上挙げてから、仮説を検証するための事例を生成、または選択した。参加者は、毎試行、仮説を口頭で報告するよう求められ、実験者が報告された内容を入力し、記録した。つづいて、選択された事例が発見すべき規則に合致するかどうかのフィードバックが与えられ、参加者はそれに基づいて、必要に応じて仮説を修正することが求められた。以上の手続きは、参加者が正しい規則を発見したと確信できるまで続けられた。

実験参加者

実験の目的を知らない、札幌市内在住の大学学部生75名(男性41名、女性34名)が実験に参加し、2-4-6問題(40名)およびエリュース課題(35名)のどちらか一方を解いた。実験は個別に実施され、参加者は、実験終了後に、それぞれの課題について知っていたかどうかを尋ねられた。いずれの参加者も、本実験で使用した規則発見課題について、事前知識がないことが確認さ

れた。

結 果

正解率および試行数

各課題で正しい規則を報告した参加者は、2-4-6 問題で 29 名 (72.5%), エリュースス課題では 19 名 (54.3%) であり、課題間で正解率に差はなかった ($\chi^2(1, N=75) = 2.69, p=.10$)。2-4-6 問題の正解到達者は、平均して 12.5 試行で規則を報告し、正解未到達者は 9.5 試行で報告を行った。エリュースス課題では、正解到達者は、12.6 試行、正解未到達者は、9.4 試行で報告を行った。課題の種類、および最終報告の正誤を独立変数、試行数を従属変数とした分散分析を行った結果、最終報告の正誤の主効果のみが有意であり ($F(1, 71) = 5.13, p=.03$)、課題の種類を含む効果は有意ではなかった。本実験の 2-4-6 問題における正解率は、過去の研究 (例えば、Wason, 1960) に比べて高くなっているが、Wason の実験参加者の正解率が低い原因として、規則報告が早すぎるものが指摘されている (Klayman & Ha, 1989)。Klayman & Ha は、全 18 試行中の 6 試行目、12 試行目、18 試行目のそれぞれで、参加者が規則と一致する仮説を挙げているかどうかを検討し、6 試行目では 22% の参加者の仮説が規則と一致するにとどまっていたが、12 試行目ではその比率が 41%、18 試行目では 52% に増加することを示した。本実験での平均試行数が、正解到達者では両課題で概ね 13 試行、正解未到達者では 10 試行であることから、本実験の正解未到達者も、十分な検証を行う前に報告を行ってしまったと考えられる。また、本実験では、正しい規則を発見したと確信できるまで検証を続けることが可能であったことも正解率の上昇に寄与していると考えられる³⁾。

事例の分類基準

本研究では、参加者が検証した事例を、Klayman & Ha (1989) に準じて、以下の基準に従って分類した。

正事例検証 (以下、+H): 仮説が 1 つだけ挙げられており、かつ事例が仮説に合致する場合。

負事例検証 (以下、-H): 仮説が 1 つだけ挙げられており、かつ事例が仮説に合致しない場合。

診断テスト (alternative test, 以下、ALT): 仮説が複数挙げられており、事例がある仮説の正事例となっているが、別の仮説に対しては負事例となっている場合。

非診断テスト (failed alternative test, 以下、FA): 仮説が複数挙げられているが、すべての仮説に同時に合致する、またはすべての仮説に合致しない場合。

以上の分類のほかに、“わからない” など、特定の仮

Table 1
Mean percentages of tests by task content

Task content	Type of test			
	+H	-H	ALT	FA
2-4-6 ^a	29.2	27.2	26.8	12.4
New Eleusis ^b	40.5	24.0	12.5	16.8
Total	34.8	25.6	19.6	14.6

^a $n=40$, ^b $n=35$.

説を挙げずに、事例の検証のみを行った場合について、Tukey (1986) は、“発見” という分類を行っているが、該当するケースが少ない (6% 未満) ため、分析の対象からは除外した。

事例選択の全体的分布

分析に先立ち、実験参加者が各試行で行った検証を、+H, -H, ALT, FA のいずれかに分類した。参加者ごとに、仮説検証過程全体での選択率を算出し、課題別の平均を示したのが、Table 1 である。先行研究と同様に、+H の選択が、他のタイプの事例に比べて相対的に多くなっているものの、その選択率自体はそれほど高いとは言えない。一方で、-H の選択率は、先行研究と比べて高くなっている。例えば、Klayman & Ha (1989) では、+H は 11.8 試行 (66%)、ALT は 3.8 試行 (21%)、-H は 1.3 試行 (7%) で選択された。また、Tweney et al. (1980) の実験 2 では、+H が 6.4 試行、-H は 3.5 試行で選択された。Tweney らの実験では、手続き上の違いにより事例が ALT に分類されていないが、本実験における ALT の全体的な選択率は、Klayman & Ha の実験と同程度であったといえる。

ここで、+H, -H, ALT の選択率の逆正弦変換値について、課題間の比較を行ったところ、+H についてはエリュースス課題のほうが 2-4-6 問題よりも選択率が有意に高く、ALT についてはその逆の結果であった (それぞれ、 $ts(73) = -2.09, 3.15, ps = .04, .002$) が、-H については、課題間の差は認められなかった ($t(73) < 1, n.s.$)。

検証過程を通じた事例選択の変化

次に、1 試行目から 5 試行目までの序盤ブロック、規則が報告される直前 5 試行の終盤ブロックにおける 4 種の事例の選択率を算出した。ここで、75 名の参加者のうち 29 名は 10 試行未満で実験を終了しており、これらの参加者については試行数の少なさから 5 試行ずつのブロックに分割できないため、以後の分析の対象からは除外した。結果として、2-4-6 問題、エリュースス課題ともに 23 名ずつが分析の対象となった。

各ブロックにおける 4 種の事例の選択率を示したのが Table 2 である。先行研究と同様に、本実験の参加

3) 最終報告に 18 試行以上を要した参加者は、2-4-6 問題では 7 名、Eleusis 課題では 3 名であった。

Table 2
Mean percentages of tests by task content and trial block

Task content	Trial block							
	First 5				Last 5			
	Type of test				Type of test			
	+H	-H	ALT	FA	+H	-H	ALT	FA
2-4-6 ^a	21.7	18.3	29.6	23.5	40.0	33.9	18.3	5.2
New Eleusis ^a	29.6	13.0	19.1	29.6	49.6	33.0	2.6	10.4
Total	25.7	15.7	24.3	26.5	44.8	33.5	10.4	7.8

^a $n=23$.

者においても、+Hは検証の終盤で増加するが、ALTは終盤において減少している。また、-Hも、+Hと同様に終盤において増加していた。ここで、+H、-H、ALTの選択率の逆正弦変換値を従属変数とし、課題の種類を被験者間要因、試行ブロックを被験者内要因とする2要因混合計画分散分析を実施した。その結果、+H、-Hについては、終盤のほうが序盤よりも選択率が有意に高かったが（それぞれ、 $F_s(1,44)=12.1, 23.61, p=.001, p<.001$ ）、課題の主効果および交互作用は有意ではなかった。ALTについては、終盤のほうが序盤よりも有意に選択率が低かっただけでなく、2-4-6問題のほうが、エリュース課題よりも選択率が高かった（それぞれ、 $F_s(1,44)=11.72, 6.55, p_s=.001, .01$ ）が、2要因の交互作用は有意ではなかった。

検証方略と課題解決の関係

次に、検証の選択と課題解決成績との関係について検討する。なお、分析の対象となった46名（両課題とも23名ずつ）のうち、2-4-6問題では17名、エリュース課題では16名が、正解に到達した。ここでは、Vartanian et al. (2003)を参考に、課題解決の成功、不成功を基準変数に、序盤・終盤における+H、-H、ALTの選択率、および課題の種類を説明変数としたロジスティック回帰分析を行った。モデルの選択は、ステップワイズ法（変数増加法）を用い、尤度比検定量の変化が有意でなくなるまで順次モデルに変数を投入した。また、課題の種類によって検証の選択が異なる傾向にあることから、課題の種類を含む交互作用項を含めて検討を行った。その結果、序盤のALTの選択率（ $\chi^2(1)=4.86, p=.03$ ）、および課題×終盤の-Hの選択率の交互作用項（ $\chi^2(1)=9.43, p=.002$ ）の2要因を含むモデル（ $R^2=.53$ ）が選択され、それ以外の変数の効果は有意ではなかったためモデルから除去された。

以上の結果から、ALTおよび-Hの使い方が、解決成績に影響を与えていることが示唆される。ここで、ALT、-Hの選択について課題、ブロック、報告の正誤

を要因とした3要因の分散分析を行った。その結果、ALTでは、ブロックの主効果が有意であり、序盤（24.3%）のほうが終盤（10.4%）よりも多く選択されていた（ $F(1, 42)=6.38, p=.02$ ）が、これ以外の主効果、交互作用はすべて有意ではなかった。-Hでは、ブロックの主効果、正誤の主効果、ブロックと正誤の交互作用、課題と正誤の交互作用が認められ（それぞれ、 $F_s(1, 42)=12.08, 7.87, 6.92, 6.69, p_s=.001, .01, .01, .01$ ）、これ以外の主効果、交互作用は有意ではなかった。交互作用が有意であったため、単純主効果の検定をしたところ、ブロック×正誤では、正解到達者では、終盤（38.4%）の方が序盤（17.8%）よりも-Hが多く選択されていたが（ $F(1, 42)=33.09, p<.001$ ）、未到達者では、序盤（11.9%）と終盤（14.8%）での選択に差は認められなかった（ $F(1, 42)<1, n.s.$ ）。課題×正誤では、エリュース課題において正解到達者（30.3%）のほうが未到達者（1.9%）よりも多くの-Hを選択していたが（ $F(1, 42)=15.25, p<.001$ ）、2-4-6問題では到達者（26.0%）と未到達者（24.8%）の間に差は認められなかった（ $F(1, 42)<1, n.s.$ ）。

10 試行未満で終了した参加者の事例選択

全参加者75名のうち29名の参加者は、10試行未満で課題を終了したため、上記の分析の対象とはなっていないが、これらの参加者の事例選択傾向についても言及しておく必要があるだろう。29名の内訳は、2-4-6問題17名（正解到達12名、未到達5名）、エリュース課題12名（正解到達3名、未到達9名）であった。

これらの参加者については、序盤、終盤に分割することは不可能であるため、+H、-H、ALTの全体的な選択率および課題種類を説明変数とし、最終的な正誤を基準変数にとったロジスティック回帰分析を試みた。その結果、-Hの選択率（ $\chi^2(1)=5.94, p=.02$ ）のみを含むモデルが有意であった（ $R^2=.12$ ）。人数が少なく課題、報告の正誤を要因とした分析を行うのは適切ではないと考えられるため、以下に-Hの選択率の平均値のみを示す。2-4-6問題においては、正解到達者が32.8%、未

到達者が18.8%, エリューシス課題においては, 正解到達者が36.7%, 未到達者が27.1%であり, 正解到達者の方がHの選択が多くなる傾向にあった。

考 察

仮説検証に関する初期の先行研究の多くは, 仮説検証方略を全体的な傾向として分析し, 正事例検証の選択が多い, すなわち確証バイアスが見られることを示してきた。しかし, 近年の研究では, われわれの仮説検証事例の選択は固定的ではないことが明らかになっている (Klayman & Ha, 1989; Tweney et al., 1981; Tukey, 1986)。本研究では, 複数の仮説群からもっともらしい仮説を抽出することを目的とした診断テストの選択は, 検証の終盤において減少するが, 正事例検証, および負事例検証の選択は増加することが示された。この結果は, Klayman & Ha (1989)をはじめとする近年の研究と一致しており, 序盤と終盤で検証選択傾向が異なるという予測が支持されたといえる。

しかし, 本研究の結果は, 先行研究と異なる点もある。まず, 本実験の参加者は, 先行研究に比べて, 負事例検証の選択が全体的に多いという点は指摘しておく必要があるだろう。この点について, 本研究の結果から明確な結論を得ることはできないが, 実験課題そのものに関する知識はなくとも, 論理学や科学哲学等の関連領域の一般知識を有していた参加者が多かったなど, 参加者集団の特徴によるものであるという可能性はある。また, 本研究では, 負事例検証の選択が, 課題解決の成功に影響するという結果が得られた。特に, 正解到達者のほうが, 終盤の5試行における負事例検証の選択率が高かった。また, 診断テストの選択については, 解決に10試行以上を要した参加者においてのみ, 課題解決の成功に影響するという結果が得られた。正解到達者のほうが負事例検証を多く選択することは, 一部の先行研究において示されているものの (例えば, Vartanian et al., 2003; Wason, 1960), これらの研究では, 最も可能性が高いと思う仮説を1つだけ挙げさせているため, 診断テストという分類カテゴリ自体が設けられておらず, 本研究の結果と直接の比較はできない。本研究でこのような結果が得られた原因については, 推測の域を出ないが, 診断テストが序盤において, 仮説の取捨選択のために用いられるのに対し, 負事例検証は終盤において, 仮説を精緻化するために用いられるものであり, 終盤で負事例検証を行い, 仮説の不備を検討することが, 課題解決を促進するのではないかと考えられる。

さらに, 本研究では, 2-4-6問題, およびエリューシス課題という2つの規則発見課題を用いたが, 2-4-6問題の方がエリューシス課題よりも診断テストの選択が多いなど, 課題により検証の選択傾向が異なるという結

果が得られた。これは, 両課題の仮説空間, 事例空間の大きさや質の違いに起因すると推測されるが, 今後は, このような課題の構造的特徴の違いを考慮に入れるとともに, 仮説生成過程も含めた, 包括的な仮説検証過程の認知モデルを構築することが必要となるであろう。

近年, Oaksford & Chater (1994) や Cherubini, Castelvécchio, & Cherubini (2005) は, それまでの仮説検証過程に関する研究が, 生成された仮説をどのように検証するかという“正当化の文脈”(context of justification)のみを扱っており, 仮説がどのように生成されるかという“発見の文脈”(context of discovery)については言及していないという問題を指摘している。本研究も, 彼らの指摘に従えば, 正当化の文脈からの分析にとどまっていると言えよう。今後は, 発見の文脈も視野に入れることで, より詳細な仮説検証過程の記述を行うことが必要であると考えられる。この点について, Tukey (1986) は, Popper (1956) の反証主義以外のモデルとして, J. S. ミルの帰納法, Lakatos (1978) のリサーチプログラム, ベイズ理論的な主観確率の変化などの概念が, 仮説検証過程の記述に有効であることを指摘している。特に, 従来の, 正事例検証 (確証), 負事例検証 (反証), 診断テスト (棄却) という分類に加え, 探索, 評価, 発見などの分類規準を提唱している点は興味深い。このうち, “発見”は, “アイディアが尽きたため, ランダムに数字列を試す”というものであるが, 本研究では, 仮説を立てずに事例を検証した試行は, その頻度が少ないこともあり分析の対象からは除外した。しかし, これらの試行では, “発見”を目的とした検証が行われたと解釈することができる。さらに, Oaksford & Chater (1994) は, 規則発見課題における有効な検証方略として, それまでに得られた事例情報を最大限に活用し, 仮説とそれに対する対立仮説を同時に検証する反復的反実仮想方略 (iterative counterfactual strategy) を挙げている。この方略も, 仮説生成のメカニズムを重視したものであるといえる。

ただし, 仮説生成過程の研究では, 以下の問題に留意する必要があるだろう。本研究を含むこれまでの仮説検証研究では, 検証を分類する都合上, 仮説を口頭や筆記により報告させていた。しかし, この手続きが, 参加者に仮説を意識化するよう促し, “発見”に分類される事例の数を減少させる効果を持つかもしれない。また, 参加者が自身の保持しているすべての仮説を正確に報告できるとは限らず, 言語化が困難なために未報告であった仮説が分析の対象とならずに, 本来, 診断テストに分類されるべき事例が, 正事例検証や負事例検証として分類されることもありうる。今後の発展的研究では, 参加者が保持している仮説を, より自然な形で正確に把握するための方法論的な改善を行ったうえで, 仮説が生成され

るプロセスをも視野に入れた包括的な枠組みから、検証過程のダイナミクスを明らかにすることが必要となるだろう。

引用文献

- Cherubini, P., Castelvécchio, E., & Cherubini, A. M. (2005). Generation of hypotheses in Wason's 2-4-6 task: An information theory approach. *The Quarterly Journal of Experimental Psychology*, **58A**, 309-332.
- Freedman, E. G. (1992). Scientific induction: Individual versus group processes and multiple hypotheses. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 183-188.
- Gardner, M. (1977). On playing New Eleusis, the game that simulates the search for truth. *Scientific American*, **237**(4), 18-25.
- Gorman, M. E., Stafford, A., & Gorman, M. E. (1987). Disconfirmation and dual hypothesis on a more difficult version of Wason's 2-4-6 task. *The Quarterly Journal of Experimental Psychology*, **39A**, 1-28.
- 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.
- Lakatos, I. (1978). *The methodology of scientific research programmes*. New York: Cambridge University Press.
(ラカトシュ, I. 村上陽一郎・井山弘幸・小林傳司・横山輝男 (共訳) (1986). 方法の擁護—科学的研究プログラムの方法論 新曜社)
- 三輪和久 (1999). 科学的発見における有効な仮説検証方略—計算機シミュレーションに基づく検討. 岡田 猛・田村 均・戸田山和久・三輪和久 (編著) 科学を考える—人工知能からカルチュラル・スタディーズまで 14 の視点 北大路書房 pp. 96-119.
(Miwa, K.)
- Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, **29**, 85-95.
- Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1978). Consequences of confirmation and disconfirmation in a simulated research environment. *Quarterly Journal of Experimental Psychology*, **30**, 395-406.
- Oaksford, M., & Chater, N. (1994). Another look at eliminative and enumerative behaviour in a conceptual task. *European Journal of Cognitive Psychology*, **6**, 149-169.
- Platt, J. R. (1964). Strong inference. *Science*, **146**, 347-353.
- Popper, K. (1959). *The logic of scientific discovery*. New York: Basic Books.
(ポパー, K. R. 大内義一・森 博 (訳) (1971-1972). 科学的発見の論理 (上・下) 恒星社厚生閣)
- Snyder, M., & Swann, W. B. J. (1978). Hypothesis testing processes in social interaction. *Journal of Personality and Social Psychology*, **36**, 1202-1212.
- 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.
- Tweney, R. D., Doherty, M. E., Warner, W. J., Pliske, D. B., Mynatt, C. R., Gross, A., & Arkkelin, D. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology*, **32**, 109-124.
- Vartanian, O., Martindale, C., & Kwiatkowski, J. (2003). Creativity and inductive reasoning: The relationship between divergent thinking and performance on Wason's 2-4-6 task. *The Quarterly Journal of Experimental Psychology*, **56A**, 641-655.
- Wason, P. C. (1960). On the failure to eliminate hypothesis in a conceptual task. *Quarterly Journal of Experimental Psychology*, **12**, 129-140.
- Wharton, C., Cheng, P., & Wickens, T. (1993). Hypothesis-testing strategies: Why two goals are better than one? *Quarterly Journal of Experimental Psychology*, **46A**, 743-758.
(2007 年 7 月 12 日受稿, 2008 年 1 月 7 日受理)