

条件文推論の学習過程

論理的推論学習支援システムに向けての実験的研究†

服部雅史*・中川正宣**

立命館大学文学部*

東京工業大学大学院社会理工学研究科**

HATTORI and NAKAGAWA (1996) および服部 (2000) は、人間の論理的な推論過程の同定のための実験的方法を提案した。本研究では、この方法が論理的推論、特に条件文推論の学習支援、および学習過程同定のための方法として利用可能であることを論じ、その有効性を確かめるため実験を行った。実験の結果、実験群においてのみ正答率の上昇が見られ、条件文推論の学習支援システムとしての有効性が示された。また同時に、被験者の発話データと対応づけることにより、学習過程における多様な特徴を同定することができた。その中で、学習過程における波及効果、パフォーマンスの劇的な変化、学習効果表出までのタイムラグ、学習の逆効果、局所的規則による見かけ上の整合性の維持などについて考察した。

キーワード：条件文推論、演繹的推論、数学教育、学習支援システム、問題解決、一階述語論理

1. はじめに

論理的推論が必要とされている場面は、科学的な推論や、裁判における推論などという特殊な場面に限られていると思われがちであるが、実際には、日常生活においても、例えば、感情的にならず冷静にものごとを考える、支離滅裂にならず筋道立てて考える、などというのはいわゆる論理的推論を指し、このような能力が必要とされる場面は極めて多い。したがって、誤った推論を正したり論理的に妥当な推論を教えたりすることは、教育的観点からいえば、単なる数学教育に

とどまらず、まさに一般教育として重要なテーマであると考えられる。

一方、HATTORI and NAKAGAWA (1996) および服部 (2000) は、論理的な問題解決事態における人間の推論過程を客観的な形式で同定する方法を提案した。現在のところ、人間の論理的推論過程を同定する実験的手法として確立されたものは、発話思考法 (ERICSSON and SIMON 1984) 以外に見当たらないが、その方法の欠如自体が、人間の推論過程、ひいてはその学習過程に関する研究をも阻んできたと考えられる。例えば、メンタルモデル理論 (JOHNSON-LAIRD and BYRNE 1991) によれば、人間の条件文推論過程の多様性 (学習による変化や個人差を含む) は、潜在的モデルを内包する初期モデルとその顕在化のしかたによって説明されるが、理論の真偽を確認するための客観的データは十分とはいえない。また、論理的推論学習過程のモデルとして有力なものは、現在のところあまりないが、今後のモデル構築のためにも基礎データは必須となる。

服部 (2000) の方法は、被験者の発話データと相補的に利用可能な実験的手法であり、この方法は、推論の正誤が被験者にフィードバックされるため、論理的推論の学習支援ソフトウェアとしての利用可能性を持

2000 年 4 月 12 日受理

† Masasi HATTORI* and Masanori NAKAGAWA** : Learning Processes of Conditional Reasoning: An Experimental Investigation towards a Learning Support System for Logical Reasoning

* College of Letters, Ritsumeikan University, 56-1, Kitamachi, Toji-in, Kita-ku, Kyoto, 603-8577 Japan

** The Graduate School of Decision Science and Technology, Tokyo Institute of Technology, 2-12-1, O-okayama, Meguro-ku, Tokyo, 152-8552 Japan

っている。もし学習への応用が可能になれば、本来の特徴である推論過程の同定と同時に学習過程の同定も可能となり、論理的推論の学習過程のきめ細かな分析が期待できる。そこで、本研究では、特に、条件文推論の学習過程の分析を目的として実験を行い、その有用性を確認した。

なお、形式論理学の学習のためのソフトウェアとしては、既に優れたものが提案されているが（例えば、BARWISE and ETCHEMENDY 1994）、本システムは、より実践的な論理的推論の習得を目指すものとして位置づけられる。

2. 推論過程同定の一般的方法

前述の HATTORI and NAKAGAWA (1996) および服部 (2000) の方法は、被験者がコンピュータ上で実施する論理的な問題解決の実験における被験者の論理的推論の過程を、演繹木 (deduction tree; 例えば CHANG and LEE 1973 参照) として同定するというものである。これらの方法の基本的アイデアは、推論ステップを論理的問題解決事態におけるオペレータ (NEWELL and SIMON 1972) に対応づけることにより、推論過程をオペレータ適用系列として獲得するというものである。以下では、服部 (2000) の方法を簡単に説明する。

被験者に与えられる課題は、コンピュータ上で実現された化学の模範実験という設定で、目標は、与えられた複数 (服部 (2000) では 8 種類) の薬品の中から指定された性質のものをすべて見つけ出すことであった。まず、被験者には、図 1 に示すような、薬品の性質の間に成り立っている「規則」を書いた紙が与えられた。規則は、架空の薬品の性質について、性質相互の (論理的) 関係を表していた。規則は一階述語論理の節集合 (set of clauses) に、薬品の性質は述語に、薬品は論理定項に、それぞれ対応づけられた。

また、薬品の性質を調べるための手段として「テスト」が用意された。被験者は、どの薬品に対してどの

テストを実行するかを、コンピュータ画面上でメニュー形式により選択した。いつ、どのようなテストを実行するかは被験者の自由であるが、予め実験者の意図により、薬品毎に実行不可能のテストが定められた (被験者は、実行時に実行可能性を知る)。さらに、薬品の性質は、ある仮想的機関に「認定申請」をし、認定されて初めて意味を持つとされた。認定申請の際、被験者は、(1) どのような結論 (薬品の性質) が、(2) どの事実 (テスト結果) とどの規則から導くことができると思うかについて、メニュー形式によって対話的に回答した。実際のテスト結果以外で認定されるのは、論理的に正しい推論結果だけとされた。こうして、課題実行中のテストと認定申請という 2 種類のオペレータの実行系列から被験者の推論過程を特定することが可能となった。

ここで導入された「認定」とは、推論ステップの同定手段であると同時に、推論の論理的妥当性を被験者にフィードバックするための手段にもなっている。つまり、被験者に対して質問や提案を発するなどの積極的な示唆は行わない代わりに、被験者が正しいと信じていた推論結果が認定されないという事実によって、被験者が自らの誤推論に気づくことを期待するしくみである。適度な難易度の課題においては、このフィードバック機構が、被験者の論理的推論学習に対して有効に機能することが期待できると考えた。

3. 実験方法

本研究では、特に条件文推論の学習に狙いを定めて課題を設計した。条件文の推論に関しては、「もし P ならば Q 」という条件文があるとき、「 P でない」から「 Q でない」を導く **前件否定の誤謬** (fallacy of denial of the antecedent) や、「 Q である」から「 P である」を導く **後件肯定の誤謬** (fallacy of affirmation of the consequent) がしばしば認められることが知られている (例えば、高橋・服部 1996, MANKTELOW 1999 などを参照)。これらの誤謬に関する効果的な学習を目的として実験を行った。

3.1. 課題

条件文推論の学習のために化学課題 I と化学課題 II の 2 つの課題を用意した。化学課題 I は、「もし P ならば Q 」の形式の条件文 (**タイプ 1** と呼ぶ) に関する正しい推論の学習を意図して設計された。化学課題 II は、「もし P かつ Q ならば R 」の形式 (**タイプ 2** と呼ぶ) に照準を当てた。また、学習の効果を見るた

1. すべての薬品は、イソ族ならば、劇性である。
 2. すべての薬品は、劇性ならば、トリ性である。
 3. すべての薬品は、トリ性ならば、重化性でない。
 4. すべての薬品は、硬類でないならば、バラ性である。
 5. すべての薬品は、硬類ならば、酸性でない。
 - 以上の規則を利用して、「硬類でなく、かつ劇性でない」薬品を全て見つけ出して下さい。

図 1 化学課題 I の「規則」

めに**文章課題**を用意した。以下では各課題について説明する。

化学課題 I 以下のような6つの論理式を用意し、(I-1)～(I-5)を「規則」、(I-6)を目標の性質とした。また、 A, B, C, D, E, F, G を、それぞれイソ族、重化性、パラ性、壊性、劇性、トリ性、硬類という架空の性質名で表現することにより、図1に示すような規則が作成された。

$$\forall x[A(x) \rightarrow E(x)]. \quad (I-1)$$

$$\forall x[E(x) \rightarrow F(x)]. \quad (I-2)$$

$$\forall x[F(x) \rightarrow \neg B(x)]. \quad (I-3)$$

$$\forall x[\neg G(x) \rightarrow C(x)]. \quad (I-4)$$

$$\forall x[G(x) \rightarrow \neg D(x)]. \quad (I-5)$$

$$\exists x[\neg E(x) \wedge \neg G(x)]. \quad (I-6)$$

実際には、上の規則の中で課題の目標達成のために必要なのは、規則 (I-2), (I-3), (I-5) だけである。いずれも**否定式** (modus tollence) を用いた推論が要求されるようになっている。否定式とは、「もし P ならば Q 」と「 Q でない」から「 P でない」を導く推論形式である。残りの2つの規則は、課題の目標達成のためには不要なダミー規則で、被験者の潜在的な誤推論を顕在化させることを意図して挿入された。具体的には、(I-1)は前件否定の誤謬、(I-4)は後件肯定の誤謬を顕在化させることを意図した。例えば、前件否定の誤謬を起こしやすい被験者は、ある個体が A でないことがわかっているとき、規則 (I-1) を使って、それが E でないと思うかも知れない。もし被験者がそのように推論して認定申請を行うと、被験者の誤推論が同定されると同時に、推論の誤りが被験者にフィードバックされるというしくみになっていた。

なお、論理定項としての薬品は全部で14個与えられ、それぞれには架空の薬品名が付けられた。この薬品数は、次に述べる化学課題 II でも同数であった。

化学課題 II 以下のような7つの論理式を用意し、(II-1)～(II-6)を「規則」、(II-7)を目標の性質とした。実際には、 A, B, C, D, E, F, G, H を、それぞれ硫化性、錯類、芳息類、遷移族、融性、リボ族、メタ族、湿性という架空の性質名で表現して規則が作成された。

$$\forall x[A(x) \wedge B(x) \rightarrow C(x)]. \quad (II-1)$$

$$\forall x[A(x) \rightarrow \neg D(x)]. \quad (II-2)$$

$$\forall x[\neg C(x) \rightarrow \neg G(x)]. \quad (II-3)$$

$$\forall x[G(x) \rightarrow B(x)]. \quad (II-4)$$

$$\forall x[D(x) \wedge E(x) \rightarrow F(x)]. \quad (II-5)$$

$$\forall x[D(x) \rightarrow \neg H(x)]. \quad (II-6)$$

$$\exists x[\neg D(x) \wedge \neg G(x)]. \quad (II-7)$$

この課題で、目標達成のために実際に用いられるべき規則は、(II-6)と(II-5)のうちのどちらか一方、および(II-4)の2つだけで、残りは、被験者の誤推論を顕在化させることを意図したダミー規則である。うち、(II-2)および(II-3)は、 A や $\neg C$ を満たす個体を意図的に用意しなかったために、実質的にダミー化した規則である。(II-2)は、 A に注目させるための規則で、 A に注目することによって(II-1)から A を導くような誤推論を顕在化する可能性を高める。同様に、(II-3)は $\neg C$ に注意を引く役目を果たし、これも(II-1)に関する誤推論の顕在化を狙っている。いうまでもなく、実際には、規則(II-1)から A や $\neg C$ という結論は導出不可能で、このことがダミー規則たる所以である¹⁾。課題の目標を達成するためには、規則(II-5)を使って、「 P かつ Q ならば R 」という規則と、「 Q である」および「 R でない」という事実から、「 P でない」を導くという、いわば拡張的な否定式の推論が必要となる。

文章課題 被験者の論理的推論の能力を測るための課題として、仮言的三段論法の課題を用意した(図2、および付録A.1., A.2.参照)。まず、第1前提文として、タイプ1:「 P ならば Q 」とタイプ2:「 P かつ Q ならば R 」の2種類の論理形式のものを用意した。次に、第2前提文として、タイプ1については、(1a)「 P でない」、(1b)「 Q である」、(1c)「 Q でない」という3種類の問題(**サブ問題**と呼ぶ)を用意した。また、タイプ2のサブ問題は次の7つとした。(2a)「 P でない」、(2b)「 R である」、(2c)「 R でない」、(2d)「 P であり、かつ Q でない」、(2e)「 P でなく、かつ Q でない」、(2f)「 P であり、かつ R である」、(2g)「 P でなく、かつ R である」。

被験者は、第1、第2前提文の各組合せに対して、両前提文を真としたときにどのような結論が得られるかの回答を与えられた選択肢(付録A.1.参照)の中

『刻印があり、かつ、重いならば、そのナイフはよく切れる。』
『そのナイフは、刻印がなく、かつ、よく切れる。』

1. そのナイフは、重い。
2. そのナイフは、重くない。
3. 特に意味のある結論は得られない。
4. 上の2つの文は互いに矛盾している。
5. その他。

図2 文章課題(タイプ2)の一例

から選んだ。課題はコンピュータ上で実施され、選択はマウスで行った。課題の材料は、化学課題のように架空の薬品ではなく、現実に存在するものの名前や性質を用いた。ただし、常識的知識が推論に及ぼす影響をできるだけ抑えるために、あまりにも常識的過ぎたり、逆に常識とかけ離れ過ぎていると思われる内容は避けるようにした。また、含意以外の意味（例えば、因果や時空間的連続性など）を明らかに含むと思われる文章は控えた。

3.2. 実験計画および手続き

化学課題による論理的推論の学習効果を見るために、被験者を2つのグループに分けた。実験群は、化学課題I、IIを実施するが、統制群は、化学課題の代わりにダミー課題（コンピュータ上のゲーム）を実施した。化学課題は、誤推論がなくなるまで同一課題を繰り返し実施した。ただし、繰り返しは6回を限度とした。

被験者の負担を考慮に入れ、実験は2日間に分けて行われた。1日目は、両群とも、まず最初に文章課題（セッション1）を実施した。次に、実験群は化学課題Iを、統制群はダミー課題を実施した。1日目の最後に、両群とも文章課題（セッション2）を実施した。2日目は、両群ともまず文章課題（セッション3）を実施し、次に実験群は化学課題IIを、統制群はダミー課題を実施した。最後に、両群とも文章課題（セッション4）を実施した。

文章課題は、第1セッションでは、タイプ1とタイプ2を各4問ずつ、計8問与えられた（1問につきタイプ1はサブ問題が3問、タイプ2は7問）。その後、セッションを重ねるたびに、各2問ずつ（計4問）新しい課題を加えていき、セッション2、3、4では、それぞれ計12問、16問、20問が課された。セッション内の各問題の提示順序、各問題中のサブ問題の提示順序は、被験者毎にランダム化された。ただし、セッション内における正答率推移を比較可能にするため、各セッションを4問題ずつに分割したときに、タイプ1とタイプ2が2問ずつになるように配置した。このように各セッションを4問題ずつに分割したときの1まとまりを、以降、セグメントと呼び、例えばセッション2の3番目のセグメントをセグメント2-3のように表すことにする。セッション1、2、3、4のセグメント数は、それぞれ2、3、4、5となった。

実験者は、化学課題終了時、文章課題の各セグメントの区切り、および文章課題の終了時に、被験者に気づいたことなどをインタビュー形式で口頭で尋ねた。

会話はテープレコーダに録音された。以降では、インタビューによって得られた発話データをプロトコールと呼ぶ。

3.3. 被験者

被験者は大学学部生とした。文章課題の第1セッションにおいて正答率が90%を越えた者は、学習過程を調べる本実験の被験者として不適切と判断し、分析対象から除外した。この基準で各群の分析対象被験者数が10人に達するまで実験を続けた。結果として、除外された者も含めると実験群は14人（被験者e01～e14）、統制群は10人（c01～c10）の被験者を要した。各被験者の所属学部・学年は以下の通りである。e01：教育4、e02：文4、e03：教育4、e05：理2、e06：工2、e07：理3、e09：文3、e10：工2、e12：農3、e14：理1、c01：文3、c02：法3、c03：農3、c04：文1、c05：理2、c06：文2、c07：文2、c08：経済3、c09：経済3、c10：理1。なお、e09とc01のみが女性であった。

4. 結果および考察

4.1. 化学課題の効果

図3は、文章課題の正答率に関するセッション毎の推移を被験者別に表している。グラフの見た目の形状から、便宜的に6種類のパターンに分類した。実験群においては、1名（被験者e09）を除く全員の正答率が向上し、劇的な向上も散見された。一方、統制群においては、正答率がやや向上した群（図3C-I）としなかった群（C-II）に二分され、劇的な向上は見られなかった。以上より、論理的推論の学習に関して、化学課題の効果があったことは明らかである。

なお、化学課題Iにおいては、のべ推論ステップ数（すなわち認定申請の総数）624に対して誤推論数は133（21.3%）であり、うちダミー規則を用いたものが85（誤推論の63.9%）であった。化学課題IIにおいては、のべ推論ステップ数387中、誤推論数は47（12.1%）で、誤推論中ダミー規則を用いたものは14（29.8%）であった。また、ダミー規則による誤推論を少なくとも1回以上経験した被験者数は、化学課題Iでは10人中9人、化学課題IIでは10人中5人であった。化学課題Iにおいてダミー規則による誤推論率が高い点から、誤推論の誘導に関してダミー規則が有効に機能したことがわかる。また、ある程度の学習が進んだ後の化学課題IIにおいて、全体に対する誤推論率と同時に誤推論中のダミー規則利用率も低下して

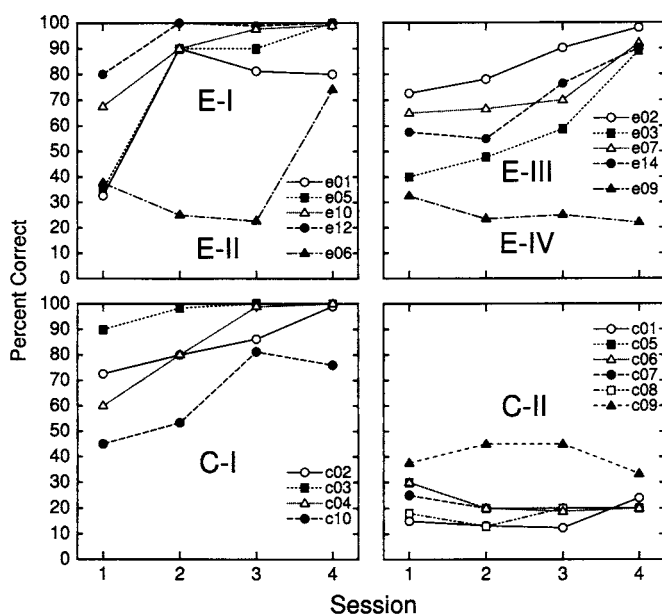


図 3 文章課題の正答率の推移。曲線の見かけの形状から E-I～E-IV (実験群) および C-I～C-II (統制群) に分類した。

いる点から、論理的推論の学習に関してダミー規則が有効であったことが推測できる。以下では、各個人の学習過程について詳細に分析・考察する。

4.2. 直接学習と波及効果

まず、被験者 e05 に注目する。この被験者は、文章課題のセッション 1 (化学課題 I を行う前) では、30% 程度の正答率であったが、化学課題 I を行った後のセッション 2 では、90% 程度にまで正答率が上がった (図 3 参照)。翌日のセッション 3 も同じ正答率で、化学課題 II の後のセッション 4 で 100% に達した。表 1 は、被験者 e05 の文章課題の各セッションにおける回答パターンとその比率の推移を示している。ここでの回答パターンは、サブ問題番号と選択肢番号 (付録 A.1. 参照) の組合わせで表されている。例えば、サブ問題 (1a) における回答として 2 番目の選択肢を選んだ場合、すなわち、「P ならば Q」という第 1 前提文と「P でない」という第 2 前提文から、「Q でない」という結論を導いた場合、その推論パターンを (1a-2) と表す。以降、推論パターンとしてこの表記を用いることにする。表 1 より、この被験者はセッション 1 の直後 (化学課題 I の後) に (2f) 以外の全問題が全問正答に変化したことがわかる。化学課題 I は 2 回繰り返す、その間に誤った推論を 6 回行った。うち (1a-2) の形式が 2 回、(1b-1) が 1 回で、残る 3 回は分類不能であった。

表 1 被験者 e05 の文章課題の推論パターンと比率 (%) の推移

Inference pattern	Session			
	1	2	3	4
1a-2	25	0	0	0
1a-3*	75	100	100	100
1b-1	100	0	0	0
1b-3*	0	100	100	100
1c-2	100	100	100	100
2a-2	100	0	0	0
2a-5*	0	100	100	100
2b-1	100	0	0	0
2b-5*	0	100	100	100
2c-4*	100	100	100	100
2d-2	75	0	0	0
2d-3*	25	100	100	100
2e-2	100	0	0	0
2e-3*	0	100	100	100
2f-1	75	100	100	0
2f-3*	25	0	0	100
2g-3*	25	100	100	100
2g-4	75	0	0	0

* 論理的正解。

以上より、セッション1と2に関して、文章課題の問題(1a)と(1b)における(1a-2)、(1b-1)という各誤推論は、化学課題Iの認定申請で誤ったことにより直接的に学習され、また、問題(2a)、(2b)、(2d)、(2e)、(2g)の誤推論の消滅は、この学習から波及的に生じた学習であることがわかる。化学課題Iの1回目終了後のインタビューにおいて、前件否定の誤謬と後件肯定の誤謬の明らかな自覚が見られ、また、文章課題のセッション2の最初の段階におけるプロトコルで、「化学実験の課題により学んだことを基にして答えた」と述べており、化学課題Iによる直接のおよび波及的学習の事実を裏付けている。同時に、タイプ2の方は「まだまだできていない」と述べ、自分が何をできないのかも把握できている。

また、化学課題IIは3回繰り返し、その間に誤った認定申請を4回行った。うち(2f-1)の形式が1回、残る3回は分類不能であった。この結果と文章課題の結果を合わせてみると、化学課題Iによって波及的学習効果が及ばなかった問題(2f)について、その後の化学課題IIによって直接的に、しかも1回で学習されたことがわかる。また、化学課題IIの直後のインタビューで、(2f-1)の形式が「必ずしも正しくない」ことを学んだと述べており、学習内容が明確に自覚されていることがわかる。

以上をまとめると、この被験者には、両化学課題が相補的に、しかも非常に効率よく、学習に役立ったことがわかる。さらに、自分が既に何を学習し、まだ何を学習していないかを自覚し、かつ明確に言語化もできており、この点は注目に値する。

4.3. 劇的なパフォーマンスの変化

図3(E-II)の被験者e06は、セッション4において急激に正答率が上昇しており、これは疑いなく化学課題IIの効果と見ることができ、この被験者は、化学課題Iでは機械のように同じパタンの反応(誤り)を繰り返すばかりで進歩がなく、結局、限度回数の6回まで実施した。その後の化学課題IIでは、1回目のある時点(全14個中12個目の薬品における性質Dの認定申請)以降、明らかにパフォーマンスが変化した。この部分の推論過程の同定結果が図4に示されている。何かをきっかけに全く新しい境地が開けるといえるのは、われわれが日常的にしばしば経験することであるが、そのパフォーマンスの変化を実験的に詳細に同定できたのはまさにこの方法の成果であり、その意義は小さくないと思われる。

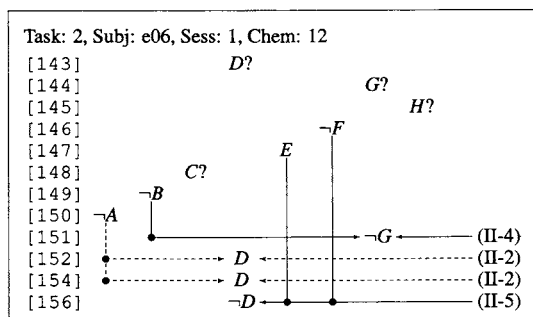


図4 被験者e06の化学課題IIの1回目の実験結果の一部。薬品番号12番の薬品についての結果のみを示す。左端の数字は、テスト実行および認定申請のステップ数を表す。また、アルファベットは薬品の性質、右側の数字は規則の番号を表す。実線は正しい認定申請を、破線は誤りを表す。例えば、ステップ143はD(遷移族)テストが実行不可能であったことを、ステップ146はF(リボ族)テストが陰性であったことを、また、ステップ151は、(II-4)の規則とステップ149の-B(錯類でない)というテスト結果から、-G(メタ族でない)という正しい認定申請(推論)を行ったことを表す。

この被験者の場合、図4のステップ152、154、156に見られるように、相矛盾する性質D(遷移族である)と-D(遷移族でない)の両方を導くことができると思ったことが直接的な変化のきっかけであったと思われる。しかし、プロトコルには、この時点での学習の自覚を示すデータはない。ただ、その後の文章課題セッション4については、「確かに(答えが)変わった」と述べているが、「どこがどう変わったかとかは説明できない」、「感覚的なもの」と説明している。

この被験者の場合、化学課題Iでは(少なくとも表面上は)全く学習が進まないまま終了した。しかし、次の段階(化学課題II)に進んだ途端、劇的な学習の進行が見られた。この結果は、1つには潜在学習による説明が可能であろう。すなわち、全く学習が進行しないように見えた化学課題Iの間も、実は潜在的に学習が進んでいたために、化学課題IIに進んで何らかのきっかけによって学習が顕在化したと考えることが可能である。しかし一方で、学習プランの不適切性を示していると捉えることもできる。すなわち、単に形式的に段階を踏むだけではなく、例えば、学習効果の表出具合と個人の学習特性などに応じてレベルを上下するような、きめの細かい学習プランが必要となる場合もあると考えられる。この点は今後の重要な課題で

表 2 文章課題のセグメント毎の正答率 (%) の推移 (実験群)

Subj.	Session 1		Session 2			Session 3				Session 4				
	1	2	1	2	3	1	2	3	4	1	2	3	4	5
e01	35	30	95	90	85	80	85	80	80	80	80	80	80	80
e02	75	70	55	95	85	80	80	100	100	100	100	95	95	100
e03	45	35	45	45	50	55	60	60	60	85	90	90	90	90
e05	40	30	90	90	90	90	90	90	90	100	100	100	100	100
e06	40	35	25	30	20	20	20	20	30	55	75	90	100	100
e07	65	65	70	65	65	70	70	70	70	75	90	95	100	100
e09	30	35	20	25	25	20	40	20	20	25	20	20	20	25
e10	65	95	100	100	100	100	95	100	100	100	100	100	100	100
e12	65	70	85	85	100	90	100	100	100	100	100	95	100	100
e14	60	55	70	45	50	45	65	95	100	90	90	90	90	90

あろう。

4.4. 学習効果表出までのタイムラグ

表 2 は、実験群の被験者の文章課題のセグメント毎の正答率の変化を示している。まず被験者 e02 に注目すると、セグメント 2-2 において正答率が急上昇し、95% に達していることがわかる。化学課題 I の後に正答率が 100% 近くに達したことから、実質的な学習は、化学課題 I においてすべて終了したと推測される。一方で、(化学課題 I の直後の) セグメント 2-1 における正答率が低いことから、化学課題 I で学習したことが文章課題に結び付くようになるまでに、ある程度の時間 (あるいはある程度の量の文章課題の実施) が必要であったという仮説が成立する。セグメント 2-1 の誤推論の内容を詳しく見ても、論理形式や問題 (内容) への依存はなく、特別の傾向はないようである。誤りに一定の傾向がないのは、学習が定着するまでの不安定な状態と捉えることができる。この被験者は、化学課題 I では同じ間違いを何度も繰り返しており、結局、最大回数の 6 回まで課題を繰り返し実施した。化学課題 I における同じ間違いの繰り返しも、学習の定着の悪さ、いわゆる「のみこみの悪さ」の 1 つの現れとして捉えることが可能である。

さらに、この被験者の場合、2 日目の最初の文章課題の前半 (セグメント 3-1, 3-2) において、やはり正答率の低下が見られる。しかも、セグメント 3-1 と 3-2 の正答率は同じであるが、詳しく見ると誤推論の内容は異なっており、やはり一定の傾向は見られない。この結果も、学習の定着の悪さに帰着できるだろう。すなわち、定着が悪かったために 1 日おいたら正答率

が低下し、再度不安定な状態に陥ったが、しばらく文章課題をやっているうちに学習内容を思い出したと捉えることができる。以上を裏付ける材料としての被験者のプロトコルが気になる点であるが、残念ながら参考になるデータは見当たらない。

さて、このような被験者の場合、どのような学習方法が効果的なのか、興味深いところである。本実験では、単純に化学課題 I を繰り返し実施したが、その間に表面化した学習効果は乏しかった。このことから、単純に化学課題 I を繰り返すというやり方は、必ずしも効果的な学習方法ではなかったかもしれない。この間の被験者の負担はかなりのものであり (化学課題 I の 6 回分の実施時間は、この被験者の場合、小休止を含めて 2 時間 30 分超)、しかも、同じ誤りを繰り返していることから、達成感が希薄であった可能性も大きい。このような被験者には、例えば、しばらく考えさせるための時間をおくとか、視点を変えた質問を投げかけるなど、何らかの別の効果的な方法が存在するかもしれない。この点も今後の興味深い課題の 1 つである。

次に、被験者 e07 の結果に着目する。表 2 の文章課題のセグメント毎の正答率の変化を見ると、セッション 4 における正答率の上昇が漸次的であることがわかる。このことから、文章課題において化学課題 II の効果が現れるためには、ある程度の時間 (あるいは、ある程度の量の文章課題の実施) が必要であったことがわかる。

学習効果表出までのタイムラグと直接は関係ないかもしれないが、この被験者には、プロトコルとパフォ

パフォーマンスのずれが見られ興味深い。まず、セグメント 4-1 の直後に、「化学実験課題により、論法を意識して答えるようになった。『 P かつ Q ならば R 』で『 P かつ Q 』の場合について変化があった」と述べている。しかし、実際のパフォーマンスを見ると、この段階ではこの形式に関して全く変化はない。次に、セグメント 4-2 の直後に、「化学実験課題を意識し過ぎて、かえって混乱し、判断に迷いが生じた」と述べている。しかし、実際のパフォーマンスは、この段階から急激に上昇している。この感想は、内的学習過程を意識化することの困難さを表していると捉えることができるかもしれない。また同時に、プロトコル（追想を含む）の信頼性の低さを実証したともいえる。プロトコルは被験者の意識的認知活動における報告であるから、それをそのまま内的学習過程の直接的な表明として捉えることには問題がないとはいえない。

最後に、被験者 e14 の正答率の変化を見ると（表 2 参照）、セッション 3 の間に段階的に急激に正答率が上昇していることがわかる。セッション 2 と 3 の間に何もなかったとすると、この上昇は、化学課題 I の効果がセッション 3 を行う間に表出してきたためと考えるのが妥当であろう。

以上、学習したことの効果が実際に表出するまでに、ある程度の時間や何らかの準備が必要であったと思われる例を見てきたが、そのしくみや理由は、各ケース毎に全く異なる可能性もある。その点に注意しながら、今後さらに事例を集めて多角的に分析していく必要があるだろう。

4.5. 混同と逆効果

被験者 e01 と e05 は、図 3 (E-I) の文章課題セッション 1, 2 に関して、正答率曲線がほぼ重なっているが、セッション 3 以降、被験者 e01 は正答率が上昇しなかった。表 3 は、被験者 e01 の文章課題のサブ問題 (2e) と (2g) に関する推論パターンとその比率の推移を示す。(2e) の正答率が全く上昇せず、(2g) の方は誤答がセッションとともに増加していることがわかる。まず、収束した誤推論 (2e-2) というのは、 $P \wedge Q \rightarrow R$ と $\neg P \wedge \neg Q$ から $\neg R$ を導くという推論パターンである。被験者は、文章課題セッション 3 終了後に、「『 P かつ Q ならば R 』と『 R でない』から『 P でないか、または Q でない』が導かれることから、『 P でなく、かつ Q でない』も含まれるのではないかと不安になってきた」と述べている。誤推論 (2e-2) の第 2 前提文と結論を入れ換えると、上のプロトコルとの

表 3 被験者 e01 の文章課題の推論パターンと比率 (%) の推移 (抜粋)

Inference pattern	Session			
	1	2	3	4
2e-2	100	67	100	100
2e-3*	0	33	0	0
2g-1	0	17	87	100
2g-2	0	17	0	0
2g-3*	25	67	13	0
2g-4	75	0	0	0

* 論理的正解。

対応が取れることから、被験者の内的過程で何らかの混同が生じていることが推測できる。次に、表 3 の (2g-1) とは、 $P \wedge Q \rightarrow R$ と $\neg P \wedge R$ から Q を導くという誤推論パターンである。これに関して、プロトコルに手がかりはないが、第 2 前提文と結論の肯定・否定をすべて逆にすると正しい推論となり、しかも、これは化学課題 II において頻繁に用いられる推論パターンであることから混同が生じた可能性が考えられる。

このようにして発生した混同を解消することに関して、化学課題 II は無力であった。一般に学習過程においては、一時的に混同が生じて不安定になることがあり得るといえる。おそらく混同が入り組んでいればいるほど、それを解消するのは困難であり、ケース毎の慎重な対応が必要となるであろう。このようなときにどのようなサポートをするかが、学習支援システムとして要求される今後の大きな課題である。

表 4 は、被験者 e03 の文章課題のサブ問題 (2c), (2d), (2f), (2g) に関する推論パターンとその比率の推移を示す。この表において、(2d), (2g) の推論の変化は、化学課題 II において直接的に体験して学習したと考えられる直接的学習、(2f) はその波及的効果と考えられる。一方、(2c) は、それまで正解していたものが、化学課題 II によって逆に誤推論に陥ったことがわかる。化学課題 II の 1 回目直後のプロトコルによると、「『 P かつ Q ならば R 』と『 R でない』とき、『 P でないか、 Q でないか』といった捉え方をしていたが、それが間違っている」と述べ、「化学課題 II の中で間違っていると言われた」と指摘している。化学課題における該当部分は、 $P \wedge Q \rightarrow R$ と $\neg P \wedge \neg R$ から $\neg Q$ を導くという誤りであり、勘違い学習をしてしまっていることがわかる。このような勘違

表 4 被験者 e03 の文章課題の推論パターンと比率 (%) の推移 (抜粋)

Inference pattern	Session			
	1	2	3	4
2c-4*	100	100	100	0
2c-5	0	0	0	100
2d-2	100	100	100	0
2d-3*	0	0	0	100
2f-1	100	100	100	0
2f-3*	0	0	0	100
2g-2	75	100	100	0
2g-3*	25	0	0	100

* 論理的正解。

いを解消するには、その学習場面を再現した上で、丁寧に間違いを指摘していく、という作業が有効かもしれない。このようなことも、学習支援システムとして望まれる機能の1つであり、今後の課題と結果といえる。

4.6. 局所的規則

被験者 e06 が化学課題 I において同じ反応パターンを繰り返して、論理的推論の学習が全く進行しなかったことは 4.3. に於いて指摘したが、この場合のプロトコルを分析してみると、興味深いことがわかる。この被験者は、化学課題 I の 4 回目の後のインタビューで、新しくわかったこととして、「この薬品 (薬品番号 12) は、イソ族から認定申請しても認定されないが、トリ性から認定申請すると認定されるということを 1, 2, 3 回目の実験で記憶していたので、イソ族のテストを実行しても時間の無駄と思い、しなかった」と述べている。つまり、規則 (I-1) を使って $\neg A$ から $\neg E$ を導く推論 (イソ族からの認定申請 = 前件否定の誤謬) と、規則 (I-2) を使って $\neg F$ から $\neg E$ を導く推論 (トリ性からの認定申請 = 否定式) の違いに全く気づいていないと思われる。そこで、イソ族とトリ性の違いは何かという実験者の間に対して、「経験上としきれない」と答えている。これら被験者のプロトコルから、論理形式上の違いを全く意識することなく、「とにかく、(理由はわからないが) この場合はこうなるのだ」という局所的な規則を作ることによって、大局的な非整合性を保持していることがわかる。局所的規則とは、特定の事例にしか当てはまらないアドホックな規則であり、日常生活において経験則や例外規則

などと呼ばれているものも含む概念である。

また、本実験において、実験群ではただ一人正答率が上昇しなかった被験者 e09 の結果も、このような局所的規則によって説明できるだろう。この被験者は、化学課題 I, II では確実に誤推論を減らして学習していることから、化学課題の領域に限定した局所的規則を作っていたという仮説を立てることが可能である。

本実験では、被験者の発言が不明確な場合を除いて、極力、実験者側から被験者に質問するのを避けたが、被験者がどのような局所的規則を作っているのかを探るためには、その点に的を絞って直接本人に質問するという方法も、場合によっては有効かもしれない。

5. 総合的考察

本研究では、HATTORI and NAKAGAWA (1996) および服部 (2000) による論理的推論過程の同定方法を、条件文推論の学習支援ソフトウェアとして利用することの可能性について論じ、実験を行った。その結果、条件文推論の学習が確認され、同時に学習過程のきめ細かい分析が可能になった。まず、本研究の方法を用いたことにより、ソフトウェアによる直接学習とその波及効果が明確に区別可能となり、定量的に測定することができるようになった。その結果、多くの学習が波及的になされていることが明らかになった。このことは、形式的推論の 1 つ 1 つが、人間にとって完全に個別的な推論パターンの集まりではなく、何らかのつながりを持っていることを示唆している。さらにいえば、形式的推論が生得的に何らかの形で潜在的に組み込まれている可能性も否定できないだろう。

また、論理的推論におけるパフォーマンスが劇的に改善する様子のみならず、その変化のきっかけとなった推論ステップまで同定できた。このような同定が可能となったのは、服部 (2000) の推論過程同定の方法を用いたためである。パフォーマンスの劇的な変化の背後には、恐らく複数のプロセスの相互作用があると考えられ、本実験で同定された一例のみから確定的な結論を引き出すことは難しいが、今後の理論化のための基礎データを提供する手段として、本研究の方法の有効性が示されたといえる。

さらに、被験者の学習過程における局所的規則を発見することができた。ここでいう局所的規則は、REPAIR 理論 (BROWN and VANLEHN 1980) における「間に合わせの修復」(repair) の概念に類似している。しかし、間に合わせの修復は、同一の行き詰まり

に対して異なる対処を行う「バグの移動」を包含するという意味で、より一般性を持った対処法を指す概念であるのに対し、局所的規則は、より個別事象に結び付いた概念であるといえる。つまり局所的規則とは、いわゆる事例ベース推論の基本的規則に近いもので、同時にスキーマやヒューリスティックスの一種とも考えられる。したがって、人間は誰しも多かれ少なかれ何らかの局所的規則を保持しているものと思われる。しかし、本実験において学習効果が認められなかった被験者のみにおいて局所的規則が同定された事実は、少なくとも本研究で扱われたような論理的課題においては、局所的規則がなじみにくいことを示しているだろう。そのことは、論理というものの抽象性を考えれば納得が行く。

また、プロトコル・データがなければ局所的規則の発見はできなかったことから、発話思考法や事後報告法によるプロトコル・データと本研究の方法のデータを相補的に利用することの有用性が明らかになったといえる。一方で、プロトコルを内的学習過程の直接的表明と考えることの問題点が明らかにされた。プロトコル・データは、メタ認知を含む被験者の（意識に上った）顕在的認知に関する報告であり、このデータの信頼性は、自己のパフォーマンスに対するモニタリングの正確さと関係が深いと思われる。したがって、認知の種類や被験者の能力によって常に信頼できるとは限らないが、その証拠が示されたといえる。

課題中の「規則」を工夫することにより、服部(2000)の方法が学習支援システムとして有用となり得ることが本研究により示された。実験後の感想の中では、「面白かった」とか「頭がよくなった気がする」などというものが多かった。このことは、一般に敬遠されがちな論理的推論の学習を、コンピュータ上の問題解決課題を通して行うことにより、比較的スムーズに行うことができる可能性が示されたといえる。

学習支援の面では、被験者の混同や勘違い学習への対処に関して、本システムの不十分性も明らかになった。このことは、被験者のモデルとして、前件否定や後件肯定の誤謬という大局的かつ静的なモデルではなく、局所的規則を含む動的なモデルが必要であることを示唆しているかもしれない。学習支援システムに内包すべき学習者モデルの精緻化と、システムの「重さ」はトレードオフの関係にあることを念頭に置きつつも、今後、より柔軟な学習者モデルを取り込むことを検討していく必要があるだろう。

注

- 1) 規則 (II-1) も (II-2) や (II-3) と同様、厳密な意味ではグミール規則ではない。というのも、もし A か $\neg C$ の個体が見つければ、(II-1) から $\neg B$ が導き出され、さらに (II-4) から $\neg G$ が出ることから、(II-1) も目標達成に寄与するからである。しかし、その場合は、一方の $\neg B$ という事実だけから、(II-3) を使うだけで同じ結論が出るので、規則 (II-1) の現実的有用性はあまりないといえる。

参考文献

- BARWISE, J. and ETCHEMENDY, J.(1994) *Hyperproof*. CSLI Publications, Stanford, CA
- BROWN, J. S. and VANLEHN, K.(1980) Repair theory: A generative theory of bugs in procedural skills. *Cognit. Sci.*, 4 : 379-426
- CHANG, C.-L. and LEE, R. C.-T.(1973) *Symbolic Logic and Mechanical Theorem Proving*. Academic Press, London
- ERICSSON, K. A. and SIMON, H. A.(1984) *Protocol Analysis: Verbal Reports as Data*. The MIT Press, Cambridge, MA
- 服部雅史(2000) 問題解決としての論理的推論—条件文の方向性に関する実験的検討. 立命館教育科学研究, 16 : 21-32
- HATTORI, M. and NAKAGAWA, M.(1996) A new experimental method to identify the process of logical reasoning. *Jpn. Psychol. Res.*, 38 : 74-84
- JOHNSON-LAIRD, P. N. and BYRNE, R. M. J.(1991) *Deduction*. Lawrence Erlbaum Associates, Hillsdale, NJ
- MANKTELOW, K.(1999) *Reasoning and Thinking*. Psychology Press, Hove, UK
- NEWELL, A. and SIMON, H. A.(1972) *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, NJ
- 高橋和弘・服部雅史(1996) 演繹的推論. 市川伸一(編), 認知心理学4 思考. 東京大学出版会, 東京, pp.15-35

Summary

Hattori and Nakagawa (1996) and Hattori (2000) proposed an experimental method to identify the logical reasoning processes of human beings. This study showed that their method is not only useful for identifying the processes, but also valid for a learning support system for logical reasoning, especially for conditional inference. The experiment

examined the validity of the method. The results showed a significant increase in the proportion of correct answers in the experimental group only, and the effectiveness of the method was demonstrated. Furthermore, with reference to the subjects' verbal data, a variety of features in the learning processes were identified and examined. The discussion includes an examination of the indirect learning processes, the drastic change in performances, the time-lag before the emergence of the effects on learning, the effects of confusion and mislearning, and the maintenance of apparent consistency using individual local rules.

Key Words: CONDITIONAL INFERENCE, TRAINING IN DEDUCTIVE REASONING, EDUCATION IN MATHEMATICS, LEARNING SUPPORT SYSTEM, PROBLEM SOLVING, FIRST-ORDER PREDICATE LOGIC

(Received April 12, 2000)

A 付 録

A.1. 文章課題の論理形式

文章課題のサブ問題（第2前提文）と選択肢の論理形式を示す。(1a), (2g)などの記号は、サブ問題の論理形式の識別子である。なお、下記の選択肢以外に、全サブ問題において、タイプ1では、「特に意味のある結論は得られない(⊥)」と「その他」の2つが、タイプ2では、「特に意味のある結論は得られない(⊥)」,「矛盾している(⊥)」,「その他」の3つが、それぞれこの順番で最後に付加された。

タイプ1『もしPならばQ』($P \rightarrow Q$)

(1a) 「Pでない」($\neg P$)

1. Qである。(Q) 2. Qでない。(⊥Q)

(1b) 「Qである」(Q)

1. Pである。(P) 2. Pでない。(⊥P)

(1c) 「Qでない」($\neg Q$)

(1b)と同じ。

タイプ2『もしPかつQならばR』($P \wedge Q \rightarrow R$)

(2a) 「Pでない」($\neg P$)

1. Qでない。(⊥Q) 2. Rでない。(⊥R)
3. Qでなく、かつRでない。(⊥Q ∧ ⊥R)
4. Qでないか、Rでないか。(⊥Q ∨ ⊥R)

(2b) 「Rである」(R)

1. Pであり、かつQである。(P ∧ Q)
2. Pであるか、Qであるか。(P ∨ Q)
3. Pでなく、かつQでない。(⊥P ∧ ⊥Q)

4. Pでないか、Qでないか。(⊥P ∨ ⊥Q)

(2c) 「Rでない」($\neg R$)

(2b)と同じ。

(2d) 「Pであり、かつQでない」($P \wedge \neg Q$)

1. Rである。(R) 2. Rでない。(⊥R)

(2e) 「Pでなく、かつQでない」($\neg P \wedge \neg Q$)

(2d)と同じ。

(2f) 「Pであり、かつRである」($P \wedge R$)

1. Qである。(Q) 2. Qでない。(⊥Q)

(2g) 「Pでなく、かつRである」($\neg P \wedge R$)

(2f)と同じ。

A.2. 文章課題の第1前提文リスト

文章課題で用いられた第1前提文を示す。先頭の番号は、問題文番号である。タイプ1, 2とも、問題01~04はセッション1~4で、問題05~06はセッション2~4で、問題07~08はセッション3~4で、問題09~10はセッション4で、それぞれ使用された。

タイプ1『もしPならばQ』($P \rightarrow Q$)

- 01 『他人の事を第一に考えるならば、その人は周りから尊敬される。』
02 『クルーザーを持っているならば、その人は裕福である。』
03 『腹の赤い魚ならば、尾びれが短い。』
04 『カードの表面にPと書いてあれば、その裏面には、数字の3が書いてある。』
05 『凹凸があれば、その商品は不良品である。』
06 『本をたくさん持っていれば、その人は博識である。』
07 『ピアスをつけていれば、その人はアメリカ人である。』
08 『よく吠えるならば、その犬は弱い犬である。』
09 『毎日ハウレンソウを食べているならば、その人は力持ちである。』
10 『二重まぶたであるならば、その人は美人である。』

タイプ2『もしPかつQならばR』($P \wedge Q \rightarrow R$)

- 01 『字が細かく、かつ、厚いならば、その本はたけになる本である。』
02 『色彩豊かで、かつ、温かい料理ならば、それは美味な料理である。』
03 『刻印があり、かつ、重いならば、そのナイフはよく切れる。』
04 『一重まぶたであり、かつ、髪が白ければ、その人は中国人である。』

- | | |
|---|-------------------------------------|
| 05 『ひげをたくわえ、かつ、足の大きな人ならば、その人は、インド人である。』 | 08 『眼鏡をかけ、かつ、はげているならば、真面目な人である。』 |
| 06 『赤い体色をし、かつ、小さい魚であれば、その魚は、熱帯魚である。』 | 09 『からだが白く、かつ、足が長いならば、その鳥は水鳥である。』 |
| 07 『髪の毛が短く、かつ、太っているならば、その女性は既婚者である。』 | 10 『アルコール度数が高く、かつ、黒ければ、それはチリの酒である。』 |