繰り返し項目についての再認の正答率と 確信度評定の関連¹

高橋晃 静岡大学

Relationship between recognition judgments and confidence ratings for repeated test items

Akira Takahashi (Shizuoka University)

Eighty-nine participants performed a set of recognition judgment and confidence rating tasks twice. Half of the new items presented in the second task had already presented as old items to the participants in the first task. Analysis of the second task responses, showed a positive correlation between confidence and performance for the old items, and a negative correlation for the new items. In particular, a strong negative correlation was observed when items presented in the first task were presented as “new” in the second task. This negative correlation reflects “source monitoring error” where the participants falsely recognized the items in the first task as those presented in the second, because they were unaware of making source misattributions.

Key words: recognition judgment, confidence rating, source monitoring.

The Japanese Journal of Psychology
2008, Vol. 79, No. 5, pp. 439-445

本論文は、自分の行った判断の正しさについての確信の程度（以下、確信度評定とする）と実際のパフォーマンスの関係について検討する。中でも特に再認判断の正答率と確信度評定の関係に注目する。

人間は、強い確信を持って行う行為は正しいことが多く、確信が低い場合にはその行為は誤っている可能性があるということを経験的に知っている。一般的に、人間が何らかの判断を行う際には、その判断に対する確信とその判断の正答率は“確信度が高いほどパフォーマンスが向上する”という正の相関関係にあることが多い（Groninger, 1976; Hart, 1967; Kelly & Lindsay, 1993; Keren, 1991）。

しかし、この正の相関関係は常に成立しているわけではない。特に、目撃者証言を実験的に扱っている各種の研究においては、目撃者の確信と実際の正答の間の相関は低い（Bothwell, Deffenbacher, & Brigham, 1987; 伊東・矢野, 2005; Lipton, 1977; Loftus, 1979 西本訳 1987; Wells & Lindsay, 1985）。こうした確信度と再認成績の関係の不安定さが何に由来するのかについ

ては、先行論文によるいくつかの示唆はあるが（Busey, Tunnicliff, Loftus, & Loftus, 2000; Gigerenzer, Hoffrage, & Kleinbultig, 1991; Griffin & Tversky, 1992; 妻藤, 2004; Thompson & Mason, 1996）これまで明確な結論が出ていない。

この確信度と再認成績の関係の不安定さの原因を示唆する論文として、確信度評定と再認判断パフォーマンスの対応関係を扱った Wagenaar (1988) がある。この研究においては、再認判断と共に確信度評定を行った場合、旧項目の正答率（hit）は確信度と正の相関関係を示していたが、新項目の正答率（Correct Rejection：以下 CR とする）は確信度の高低によらず一定の値を示していた。ここから、先行研究における確信度評定と再認成績の関係の不安定さの根底にある要因が示唆されたが、こうした現象の原因自体は不明なままであった。

高橋・海保（1998）はこれを追試し、記銘のしやすさを変化させた日本語の刺激材料で現象が再現することを確認した。また高橋（1998）では、一項目あたりの学習時間を 0.5 秒に制限し、かつ音読を行わないという、高橋・海保（1998）と比較して旧項目の痕跡強度を弱く設定した条件において、新項目の正答率が確信度評定と“弱い負の相関関係”を示すことを見出した。この“確信度が高いほどパフォーマンスが低下する”という結果は、経験的知識と矛盾する現象であり、また先行研究においてもこうした結果を見出した

Correspondence concerning this article should be sent to: Akira Takahashi, Faculty of Informatics, Shizuoka University, Jo-hoku, Hamamatsu, Shizuoka 432-8011, Japan (e-mail: akirtaka@inf.shizuoka.ac.jp)

¹ 本研究の一部は、筑波大学心理学研究科および生命工学工業技術研究所在籍中に行われた。本研究の一部は、国際記憶会議第 2 回大会および日本心理学会第 63 回大会において発表された。

ものはなかった。

この高橋 (1998) では、こうした結果の原因として、旧項目の記憶痕跡を弱くした実験条件から、ソースモニタリングエラーに基づいた誤った判断が存在するものと仮定されていた。再認判断とは、その判断の性質上、処理の内部にソースモニタリングの過程を含まなければならない (Johnson, Hashtroudi, & Lindsay, 1993)。実験参加者が再認判断と確信度評定を行う際に、旧項目の場合には、ターゲット記憶痕跡の強度など、実験時エピソードに直接関連した情報を参照できる。一方、新項目判断のように、はじめから実験時エピソードが存在しない場合には、事前の項目熟知度など、実験時エピソードの記憶痕跡以外の情報源を元にして判断を行う可能性が高くなる。その際、確信度評定時において、そうした再認判断と無関係な情報源を利用して、そうした再認判断と無関係な情報源を利用していることに実験参加者自身が気づけば、そこで判断を修正できるため、確信度と正答率の間には正の相関関係が期待される。しかし、実際には確信度と正答率の関係は正の相関を示すことがなかった。したがって、この現象の背後には、実験参加者が確信度評定を行う際に、評定のための情報源を自分では把握できず、ターゲットエピソードとは無関係な情報に基づいて確信度評定を行ってしまうソースモニタリングエラーが存在すると結論付けられていた。

しかし、高橋 (1998) では、記憶痕跡が弱いという状態と、ソースモニタリングのエラーを対応づける関連性が弱いという問題があった。なぜなら、旧項目の記憶痕跡が弱いという状態からは、有効な記憶痕跡が利用できない、という消極的な状態が示唆されるにとどまり、ソースモニタリングエラーの根拠となる、その他の誤った情報源を参照しているという積極的な証拠が見いだせなかったためである。

そこで、本論文においては、確信度とパフォーマンスの関係の不安定さの原因をソースモニタリングエラーであると仮定した上で、誤った情報を参照しやすい実験環境を構成した。その場合に、記憶痕跡が強い条件においても、その誤った情報を積極的に参照することで、正答率と確信度評定が負の相関関係を示す可能性があることを予想した。

方 法

材料 林 (1976) より連想価 31—40 までの項目合計 108 項目を利用した。この中からランダムに 36 項目を第 1 課題の旧項目、同数を新項目として割り当てた。また、第 1 課題の旧項目の半数を第 2 課題の旧項目に、残りの半数を第 2 課題の新項目に割り当てた。第 2 課題の残りの項目は、それまで提示されていない残りの材料の中から選択された。

実験参加者 大学生 89 名、年齢 18 歳から 24 歳。

手続き パーソナルコンピュータを用いた個人実

験。実験は二つの課題からなる。実験参加者は実験全体の流れを教示され、まず練習として教示された第 1 課題に入った。はじめに、画面上に連続提示される学習項目 36 項目について初期学習を行った。このとき、各項目の提示時間は 2.0 秒間 (刺激間隔 0 秒) で、実験参加者は各項目を音読し、4 項目ごとに提示される項目が、直前の 4 項目に含まれているか否かについてできるだけ速く判断を行った (初期学習)。この旧項目の学習の後、3 分間の迷路課題をはさみ、新項目と旧項目を画面上に一つずつランダムに提示した最終再認判断を 72 項目に対して行った。実験参加者はまず新旧の判断を行い、次の画面でその判断の正しさについての自己評定を 50% から 100% までの 6 段階で評定した。このとき、“自分の行った判断が絶対正しいと思った場合には 100% を押し、自信がなく、ほとんどでたために行ったと感じた場合には 50% を押す”ように教示された。最低の確信度が 50% である理由 (二者択一の問題であるため、でたために判断を行った場合でも 50% の確率で正答できる) も説明された。判断はすべて“できるだけ早く行う”という教示がなされていた。第 1 課題が終了すると、5 分の休憩をはさんで、再度第 1 課題と同じ手続きの第 2 課題を行った。このとき、第 2 課題の新旧各材料の半分には第 1 課題の材料が含まれていた。第 2 課題の開始時において“さきほど行ったのは練習試行であり、これから行う試行が正式な実験である”という教示がなされた。すべての実験が終了した後、実験の目的が説明された。実験は 40 分で終了した。

結 果

第 1 課題において、旧項目と新項目の正答率の差が 40% 以上あった実験参加者は、“あった”ないし“なかった”どちらかだけの反応傾向が強いものとして分析から除いた。また確信度評定が 6 段階の評定スケールで 3 カ所以上の評定をしていることを最低基準として、その基準未満の実験参加者を“内的確信度の相違に鈍感な実験参加者”として分析から除いた。その結果、有効実験参加者数は 79 名となった。なお、分析のため各実験参加者のデータをすべてプールし、プールされたデータを各確信度評定値毎に振り分けて集計した。

Table 1 に第 1 課題および第 2 課題全体の hit 率と CR 率、および確信度評定値を示す。さらに Table 2 に、第 2 課題を条件別に区分した同様の分析、すなわち、第 2 課題で初めて出現した項目 (以下、初出項目とする) と第 1 課題ですでに出現していた項目 (以下、既出項目とする) とに分類した場合の hit 率と CR 率、ならびに確信度評定値を示す。

第 1 課題、および第 2 課題の“条件をまとめた全体”における正答率と確信度評定の対応関係、ならび

Table 1

Average proportion correct and confidence rating in task 1 and task 2 (all items)

	Proportion Correct (Confidence)	
	Task 1	Task 2
Old Items: hit	64.5% (77.4%)	70.1% (78.8%)
New Items: CR	77.1% (75.2%)	56.2% (75.3%)

Table 2

Average proportion correct and confidence rating in task 2 (each item conditions: not presented or presented in task 1)

	Proportion Correct (Confidence)	
	Not-presented in task 1	Presented in task 1
Old Items: hit	62.0% (77.5%)	78.3% (80.2%)
New Items: CR	71.6% (75.4%)	40.8% (75.2%)

に各評定の評定度数を Figure 1 に示す。第 1 課題については、正答率と確信度の対応関係については、高橋 (1998) の強痕跡条件と同様、旧項目は確信度と正答率に正の相関関係が認められ、新項目については確信度によらず、正答率は一定であった。これを確認するために正答、誤答の度数分布で 2 (正答/誤答) × 6 の χ^2 検定を行ったところ、旧項目では分布の有意差があり ($\chi^2(5, N=2772)=418.87, p<.01$)、残差分析の結果からは、全体の正答、誤答の比率と比較して、50% から 70% までの正答率が有意に低く (調整された残差 “以下同じ” 50% 評定: -11.87; 60% 評定: -9.93; 70% 評定: -3.83; いずれも $p<.01$)、また

90% 評定から 100% 評定までの正答率が有意に高かった (90% 評定: 6.72, $p<.01$; 100% 評定: 14.16, いずれも $p<.01$)。ここから、確信度の高低と正答率の高低が対応しているといえる。一方、新項目について同様の検定を行ったところ、新項目全体の正答率と比較して、各確信度における正答率の有意差は認められず ($\chi^2(5, N=2772)=7.96, ns$)、新項目については正答、誤答の比率が確信度の高低によって変化しないことが示された。第 1 課題については、基本的には高橋 (1998) の強痕跡条件と同じ実験条件であるため、新旧項目の正答率が確信度評定との関連において分離する現象をあらためて確認した形となった。

Figure 1. Relation between the confidence rating and the proportion correct in task 1 and task 2 (all items).

第 2 課題全体については、旧項目は確信度と正答率に正の相関が認められるが、新項目は確信度が高くなるほど正答率が低下する現象が認められた。これを確認するために正答と誤答の度数分布で 2 (正答/誤答) × 6 の χ^2 検定を行ったところ、旧項目では分布の有意差があり ($\chi^2(5, N=2772)=441.4, p<.01$)、残差分析の結果からは、全体の正答：誤答の比率と比較して、50% から 70% までの正答率が有意に低く (50% 評定：-12.51; 60% 評定：-9.29; 70% 評定：-5.49; いずれも $p<.01$)、また 90% 評定と 100% 評定の正答率が有意に高かった (90% 評定：5.98; 100% 評定：14.54; いずれも $p<.01$)。ここから、旧項目については確信度と正答率のあいだに正の相関関係があることがわかった。一方、新項目について同様の検定を行ったところ、新項目全体の正答率と比較して、全体的な分布に有意な偏りが認められた ($\chi^2(5, N=2772)=50.11, p<.01$)。しかし、残差分析からは、旧項目とは反対に、全体の正答、誤答の比率と比較して、低確信度の 50% 評定と 60% 評定の正答率が有意に高く (50% 評定：5.11, $p<.01$; 60% 評定：2.22, $p<.05$)、また 90% 評定と 100% 評定の正答率が有意に低

いことがわかった (90% 評定：-4.54, $p<.01$; 100% 評定：-1.98, $p<.05$)。これは高橋 (1998) で認められた、確信度と正答率のあいだに負の相関関係が示される現象を再現していた。

次に、第 2 課題の初出項目と既出項目の 2 種類を区別して同様の分析を行った。初出項目および既出項目についての正答率と確信度評定の対応関係を Figure 2 に示す。初出項目の場合には、それだけを取り出せば、条件としては第 1 課題と同じである。結果も、新項目の 100% 評定の部分が突出していたが、それ以外は基本的に第 1 課題の結果と似たものとなった。これを確認するため、正答、誤答の度数分布で 2 (正答/誤答) × 6 の χ^2 検定を行ったところ、旧項目では分布の有意差があり ($\chi^2(5, N=1386)=208.41, p<.01$)、残差分析の結果からは、全体の正答、誤答の比率と比較して、50% から 70% までの正答率が有意に低く (50% 評定：-7.66; 60% 評定：-5.85; 70% 評定：-4.27; いずれも $p<.01$)、また 90% 評定と 100% 評定の正答率が有意に高かった (90% 評定：4.04; 100% 評定：10.90; いずれも $p<.01$)。ここから、確信度の高低と正答率の高低が対応しているといえる。一方、新

Figure 2. Relation between the confidence rating and the proportion correct in task 2 (each-item condition; not presented or presented in task 1).

項目について同様の検定を行ったところ、新項目全体の正答率と比較して、分布の偏りに有意傾向が認められた ($\chi^2(5, N=1386)=9.99, .05 < p < .10$)。残差分析から、全体の正答、誤答の比率と比較して、100%評定の正答率だけが有意に高いことがわかった ($3.09, p < .01$)。

既出項目については、旧項目は従来と同様に確信度と正答率が正の相関を示していたが、新項目については確信度と正答率が負の相関を示していた。これを確認するため、まず旧項目に対して正答、誤答の度数分布で2 (正答/誤答)×6の χ^2 検定を行ったところ、分布の有意差があり ($\chi^2(5, N=1386)=229.63, p < .01$)、残差分析の結果からは、全体の正答、誤答の比率と比較して、50%から70%までの正答率が有意に低く (50%評定: -10.05 ; 60%評定: -7.24 ; 70%評定: -2.89 ; いずれも $p < .01$)、また90%評定と100%評定の正答率が有意に高かった (90%評定: 4.57 ; 100%評定: 9.25 ; いずれも $p < .01$)。ここから、確信度の高低と正答率の高低が対応しているとわかる。これに対して、新項目について同様の検定を行ったところ、新項目全体の正答率と比較して、有意な分布の偏りが認められた ($\chi^2(5, N=1386)=118.23, p < .01$)。残差分析から、全体の正答、誤答の比率と比較して、50%評定から70%評定までの正答率が有意に高く、(50%評定: $7.42, p < .01$; 60%評定: $3.47, p < .01$; 70%評定: $2.40, p < .05$)、また90%評定から100%評定までの正答率が有意に低いことがわかった (90%評定: $-5.77, p < .01$; 100%評定: $-5.44, p < .01$)。ここから、新項目については確信度と正答率の間に負の相関関係が認められた。

また、高確信度時の正答率が偶然の割合である50%を下回っていることに対して、各確信度の正答率について χ^2 検定を行ったところ、80%から100%について、正答率が偶然の割合である50%を有意に下回っていた (80%: $\chi^2(1)=21.30$; 90%: $\chi^2(1)=65.02$; 100%: $\chi^2(1)=35.02$; いずれも $p < .01$)。

考 察

結果として、第1課題においては、高橋 (1998) で認められた、新項目の正答率が確信度評定と無関係に一定になる現象があらためて確認された。また、第2課題については、第2課題全体における新項目について正答率と確信度の間に負の相関が認められ、材料を区分して分析したところ、その原因は、第1課題において一度提示され、そのエピソードとの関連を持つ既出項目にあることが確認された。

これは、第1課題において出現して、かつ第2課題においては出現していなかったはずの項目について、実験参加者が第2課題のエピソードに帰属するとして誤った再認判断を行っていることを示唆している。第

2課題においては“練習課題と本課題は別であり、本課題の出現項目にのみ注目する”という教示を行っており、実験参加者もそのことを理解して行っていたことは観察できた。しかし、その一方でこうした結果が生じたことは、第1課題と第2課題を個別の実験エピソードとして弁別することが実験参加者にとって困難であったことを示唆している。

さらに、第2課題の既出項目の新項目において、確信度が高いほど正答率が低下する現象がより明確に認められた。100%の確信度時に偶然の割合である50%にまで正答率が低下した高橋 (1998) と異なり、本実験の場合には、高い確信度評定の際に、偶然の割合を有意に下まわる正答率を示した。このことは、実験参加者が誤った手がかりを積極的に用いて判断を行っているが、その誤りに気がつかず、むしろ強い確信を持ってそうした手がかりを利用していることを示唆している。本実験の場合には、材料設定として第1課題に出現していた既出項目についてのみ、そうした現象が生じていることから、第2課題において誤って第1課題の実験時エピソードを参照しているものと考えてよい。

本実験の第1課題、ならびに Wagenaar (1988) と高橋 (1998) で示された“新項目のパフォーマンスが確信度に依存せずに正答率が一定の状態になる”という現象は、新項目の再認判断が全体としてある値のパフォーマンスを示す一方で、その確信度については、実験参加者が再認判断の正答と無関係な手がかりを用いて評価していることを示唆している。この現象は、Busey et al. (2000) による、記憶と確信度が各々異なった独立変数に依存しているという観点と一致する。ここで、第2課題の既出項目の場合には、実験条件の設定により、その誤った手がかりが、第2課題の正答と相反するエピソードになっていたために、正答率と確信度の間に負の相関が出現するという今回のような結果が得られたものと考えられる。

このことは、確信度評定の際に、旧項目を判断するための情報だけが有効な情報として用いられ、新項目を正しく判断するための手がかり情報、たとえば悉皆検索の結果が否定的であることなどは、確信度評定に利用されていないことを示唆している。このことは、新項目の有無に正しく関連した手がかり情報が確信度評定に利用されていると仮定した場合には、新項目も、旧項目と同様に確信度とパフォーマンスが正の相関を示すと考えられることから推測される。

そして、実験参加者自身に、その確信度評定の新旧項目における手がかり利用の状況の相違がモニターできないことが、確信度と正答率の関係が不定であるという本現象の本質的な原因であると考えられる。なぜなら、誤った手がかりを用いている状態について実験参加者が正しく感知することができれば、こうした現

象は発生しないためである。実験参加者の意識上では、手がかりは過去エピソードとの関連で評価されると想定されるが、本実験条件においては同じ実験を同質の材料で繰り返していたため、それがどちらの実験エピソードだったのかという識別が困難であったと考えられる。そのため、ターゲットである特定の実験エピソード以外のエピソードが侵入し、強い確信を伴う誤った判断がなされてしまった。しかし、実験参加者はこうした誤情報の侵入に対して無自覚な状態で確信度を評定していた。これがソースモニタリングエラーである。

こうしたソースモニタリングエラーは、本実験のような再認判断時に限らず、確信度評定を行う必要のある様々な認知的場面において無意識的に発生している可能性が高い。先行研究において、確信度が正確なパフォーマンスを反映しない現象の一因は、こうした要因から発生していると予想される。

本論文においては、高橋 (1998) と異なり、記憶痕跡それ自体が強い条件でも、新項目において確信度と正答率が負の相関を示す現象が認められた。したがって、新項目判断時における確信度評定に関する正答率が一定ないし負の相関を示す現象の生起は、高橋 (1998) において条件設定されていた記憶痕跡の強弱が本質的原因ではなく、確信度評定時に誤った手がかりを利用していること、およびそれを自覚できないソースモニタリングエラーに由来する現象であると結論づける。

なお、本論文ならびに Wagenaar (1988)、高橋 (1998) 等の先行論文から、再認判断時の確信度評定において、人間が新項目判断に対しての確信を適切に持てないという現象があらためて明白になった。本論文において明らかになったソースモニタリングエラーは、この現象に関連する一因であると考えられる。

しかし、なぜ新項目に対する確信度評定時においてのみ、そうしたエラーが発生するのか、あるいはなぜ旧項目に対する確信度評定は正しくパフォーマンスを反映するのか、という両項目の確信度の処理プロセスの相違の理由についての問題は残されている。この現象は、一つの実験結果の中に確信度とパフォーマンスの相関関係が高い場合と低い場合とを同時に含むことから、確信度研究において重要な注目点となりうるものと考えられる。今後の一層の検討を必要とする。

引用文献

- Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology*, **72**, 691-695.
- Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F.

- (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychological Bulletin and Review*, **7**, 26-48.
- Gigerenzer, G., Hoffrage, U., & Kleinböltig, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, **98**, 506-528.
- Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, **24**, 411-435.
- Groninger, L. (1976). Predicting recognition during storage: The capacity of the memory system to evaluate itself. *Bulletin of the Psychonomic Society*, **7**, 425-428.
- Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior*, **6**, 685-691.
- 林 貞子 (1976). ノンセンスシラブル新規準表 東海大学出版会 (Hayashi, S.)
- 伊東裕司・矢野円郁 (2005). 確信度は目撃記憶の正確さの指標となり得るか 心理学評論, **48**, 278-293.
- (Itoh, Y., & Yano, M. (2005). Can eyewitness confidence be used as an index of accuracy of memory? *Japanese Psychological Review*, **48**, 278-293.)
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.
- Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, **32**, 1-24.
- Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica*, **77**, 217-273.
- Lipton, J. P. (1977). On the psychology of eyewitness testimony. *Journal of Applied Psychology*, **62**, 90-95.
- Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.
- (ロフタス, E. F. 西本武彦 (訳) (1987). 目撃者の証言 誠信書房)
- 妻藤真彦 (2004). 確信度評定のメカニズムと理論的問題 風間書房 (Saito, M.)
- 高橋 晃 (1998). 再認の正答率と確信度評定の関連について 心理学研究, **69**, 9-14.
- (Takahashi, A. (1998). The relationship of the proportion correct in recognition and the confidence rating. *Japanese Journal of Psychology*, **69**, 9-14.)
- 高橋 晃・海保博之 (1998). 再認実験における MOC 曲線の利用についての問題 筑波大学心理学研究, **20**, 85-90.
- (Takahashi, A., & Kaiho, H. (1998). The relations between recognition memory performance and confidence rating. *Tsukuba Psychological Research*, **20**, 85-90.)
- Thompson, W. B., & Mason, S. E. (1996). Instability of individual differences in the association between

- confidence judgment and memory performance. *Memory and Cognition*, **24**, 226-234.
- Wagenaar, W. A. (1988). Calibration and the effects of knowledge and reconstruction in retrieval from memory. *Cognition*, **28**, 277-296.
- Wells, G. L., & Lindsay, R. C. L. (1985). Methodological notes on the accuracy-confidence relation in eyewitness identifications. *Journal of Applied Psychology*, **70**, 413-419.
- 2005. 2. 21 受稿, 2008. 7. 5 受理——