

論 文

再認記憶の理論における信号検出パラダイムの問題点

A problem of the signal-detection theory paradigm in recognition memory theories.

妻 藤 真 彦

これまで再認記憶の研究では、被験者が解答を決めるとき判断過程を理論化するとき、信号検出理論 (signal detection theory; e.g., Tanner & Swets, 1954; Swets, Tanner, & Birdsall, 1961) の枠組みが使われてきた (e.g., Murdock, 1974)。元々は、感覚刺激を検出する能力の理論であった信号検出理論を、記憶理論の一部として取り入る場合、再認において記憶系の出力が、何らかの1次元の量 (熟知度、既視感の程度など) に変換され、それが、ある基準値よりも強ければ「見たことがある」等と判定されるのだと仮定される。これまで、記憶課題に信号検出理論をベースにしたモデルが当てはめられる根拠は、ROC曲線 (Receiver Operating Characteristic Curve) への適合が中心であった (e.g., Bernback, 1967; Donaldson & Murdock, 1968; Kintsch, 1967; Murdock & Dufty, 1972; Norman & Wickelgren, 1965; Murdock, 1974, for review)。また、そのような判断の確信度についても、この1次元量 (基本理論ではstrength) と基準値の差の絶対値によって決まると仮定されてきた (e.g., Barakrishnam & Ratcliff, 1996; Glanzer, Adams, Iverson, & Kim, 1993; Wallsten & Gonzalez-Vallejo, 1994; 妻藤, 2000, for review)。

このように、記憶や知識に関する再認型の課題での正答率や確信度を説明する理論では、信号検出理論は、非常に重要な位置を占めている。しかし、Saito (1998) は、この内、確信度の部分について疑問を抱かせる現象を発見した。というよりも、真っ向から予測と対立するデータを示した。この論文では、記憶判

断過程自体に関する信号検出理論の枠組みは壊さずに、確信度の部分だけを修正する仮説が提案されている。ただし、この仮説は、かならずしも記憶判断系が信号検出理論タイプであるかどうかとは関係なく、確信度評定システムとして働くようなものである (妻藤, 1999)。

Saito (1998) のデータは、確信度評定について、これまでの理論の検証が不十分であり、まだ理論から予測できる現象が未検証のままに残っていることを指摘して、この予測について実験を行った結果である。しかし、記憶判断系の理論自体についても、以下で述べるように未検証の予測が残されている。しかも、Saito (1998) の実験結果の中には、確信度の理論だけではなく、この未検証である信号検出理論の性質についても、検証データとして使用できるものが含まれている。本論文の目的は、この点の検討である。

未検証の予測

信号検出理論では、2肢選択式の再認の場合、2つの選択肢に関する「強さ」の差によって解答が決まるため、判断基準値は予測において無視することが出来る (妻藤, 2000, 参照)。そこで、この「強さ」の差をグラフの横軸にとって確率変動を表現すると、Figure 1のようになる。ある特定の「強さ」の差が、特定の実験試行で出現する確率密度が縦軸である。この図 (および以下の図) では、便宜上、正答が右 (正の側) になるように描いてある (Figures 1と2では、図を見やすくする便宜上、途中でカットしているの、積分

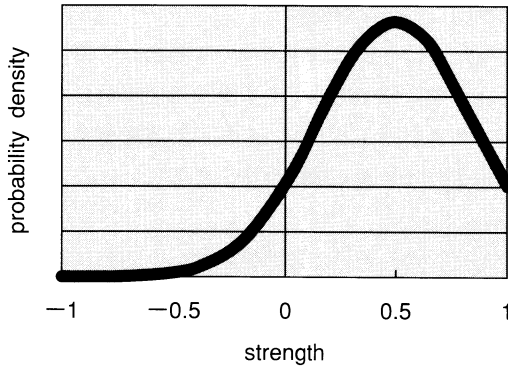


Figure 1 Probability distribution of an answer-alternative strength difference

は1ではなく、また横軸の数値スケールも理論ごとに異なる)。

Saito (1998) のような一般知識問題セットの場合、被験者の知識レベルに応じて、各々の質問項目に関する「強さ」の差は、異なった確率分布を持つはずである(質問項目・選択肢セットのヘテロジニアティ)。Figure 2に示すように、ある質問項目は、多くの被験者が良く知っており、ピークが最も右にあるような分布を示すであろう。一方被験者が全く知らない質問項目は、ちょうどゼロのところにピークがあるような分布になる。

この図には、もう一つ、ピークが左側(誤答側)にある項目も入っている。これは、ある程度の被験者が間違っ

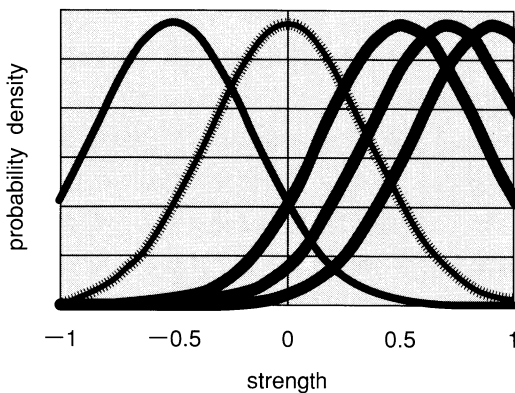


Figure 2 Heterogeneity of question-item set.

一貫して間違っ

た推測が行われるような項目である。例えば、都市名からそれが属する国名を選ぶような場合、都市名の発音が他国の言語の発音と類似しているために、多くの被験者が誤答になる選択肢を選んでしまうような場合である。Saito (1998) は、同一の質問セットを、2つの実験セッションで2回呈示した。各質問セットは、48個のヨーロッパの都市名がひとつずつ呈示され、それが当てはまる国名の選択肢が2個同時呈示されるというものであった。このようなときのセッション間での繰り返しに関する、解答の確率変動を考えてみよう。Figure 1の曲線の、ゼロより右の領域の面積(p)が、この質問項目に関する正答率になる。誤答はゼロより左側、つまり $1-p$ である。Figure 2が示すように様々なピークを持つ質問項目が呈示され、ひとつひとつ解答された後、もう一度同じ質問セットが呈示された場合に解答が変化

する確率も、このような分布から簡単に予測することが出来る。ある特定の項目について(Figure 1)、質問が繰り返されたときに解答が変化

$$C = 2p \cdot (1 - p) \quad (1)$$

する確率は、次のようになる。まず、一回目に(たまたま)右側の値が出現したなら、2回目に反対側が出現する(条件付)確率は、 $1-p$ である。そして、一回目に右側の値になる確率は p であるから、このようなことが起こるのは、 $(1-p) \cdot p$ になる。同様に、一回目が左側であったなら、解答が変化

する確率は、 $p \cdot (1-p)$ 。すると、全体として、この項目について解答変更が起こる確率(C)は、

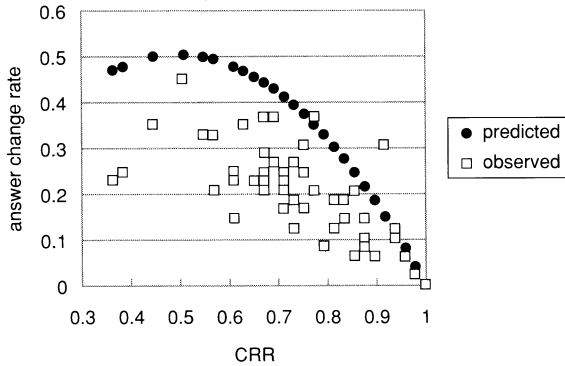


Figure 3 Observed and predicted relations between answer change rate and CRR (correct response rate) in session 1.

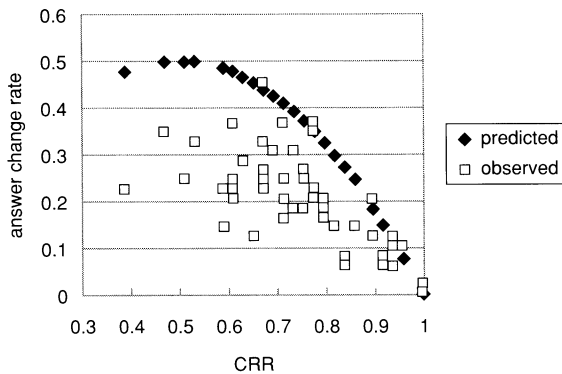


Figure 4 Observed and predicted relations between answer change rate and CRR (correct response rate) in session 2.

されている。また、同時に、式 (1) の C (解答変更率の信号検出理論による予測値) も、実測正答率ごとに示してある。これを見れば明確であるように、予測と実測は何の一致傾向も示していない。予測の項目間平均値は0.35であり、実測の項目間平均値は0.21である。項目間の分散をベースに検定すると、明確な有意差がある ($t > 10^5$, $p < .001$)。

Figure 4には、セッション2での、実測正答率との関係が示されている。これも、Figure 3の場合と同様に、明らかに予測と実測は無関係であり、予測平均 (0.34) と実測の間には同様な有意差がある ($t > 10^5$, $p < .001$)。

信号検出理論と記憶過程

感覚系の信号検出課題について幅広く有用性が示されてきた信号検出理論であるが、再認記憶にそのまま適用してよいのかどうか、今回の解答変更率に関する予測の失敗は大きな疑問を提出することになるかもしれない。

ただし、これだけで最終結論にいたるわけにはいかない。前節での検証は被験者間の相違を考慮に入れない。前節での検証は被験者間の相違を考慮に入れない。前節での検証は被験者間の相違を考慮に入れない。つまり個々の被験者の (各項目に関する) 知識の程度を無視して、ある項目が、この被験者集団においてどの程度知られているかという尺度になっている。ということは、被験者間の相違を一切考慮にいれていない計算であるため、予測と実測のズレは、これによって生じた可能性も、当然、残っている。そこで、被験者の知識レベルのばらつきによって、どのようなズレが生じるかを検討しておく必要があり、それによって、ここでの結論には、以下のような制約を設ける必要が生じる。

できるだけ簡単にこの問題を検討するために、最も極端な場合、すなわち、ある質問項目について全く異なる分布型を持つ2種類の被験者がいるものとしよう。このようなことが起こっていたとすると、解答変更率に関する予測と実測の相違は最大になる。例えば、被験者の半分は、ある質問項目について、100%完璧に知っており、間違えることも解答がふらつくこともないが、他の半分はこの項目について間違った答えを覚えてしまっているような場合である。これが最も極端なケースであり、それ以外のときは、予測と実測の相違は中間の値になる。

このような極端な場合を表現するために、あるタイプをAとし、このタイプの正答率を P_A 、もう一方をBとして、この場合の正答率を P_B とする。タイプAとBの被験者集団の中での割合を、 a および $1-a$ とすると、被験者集団全体での正答率と解答変更率の関係は、

$$C = 2 a P_A (1 - P_A) + 2 (1 - a) P_B (1 - P_B) \quad (2)$$

であるから、書き直すと、

$$C = 2 \{a P_A + (1 - a) P_B\} [1 - \{a P_A + (1 - a) P_B\}] - 2 a (1 - a) (P_A - P_B)^2 \quad (3)$$

になる。 $a P_A + (1 - a) P_B$ は、ある質問項目の被験者全体の平均正答率であり、式(1)の p に等しい。すると、式(3)の第1項は、式(1)と同じことを示している。つまり、項目の正答率と解答変更率の間に二次関数の関係がある。そこで、第2項は、被験者集団が一様ではないときの、式(1)の予測・実測差をあらわしていることになる。明らかに被験者が一様であれば($P_A = P_B$)、この予測・実測差項はゼロである。

各質問項目において、予測・実測差項の値には、 $a P_A + (1 - a) P_B$ が(実測の)正答率に等しいという制約がある。ということは、正答率が50%から離れた値になるほど、(理論的にあり得る)予測・実測差項の大きさは小さくなる。

ある項目の被験者集団全体の正答率が0.5 ($a P_A + (1 - a) P_B = 0.5$)のときに、予測・実測差が最大になるのは、 a が0.5、 $P_A - P_B$ が1.0のときである。このときの予測・実測差は、50%であるから、予測される解答変更率はゼロになる。ただし、このようになるのは P_A か P_B のどちらかが1.0で他方が0のときであり、まずあり得ない。なぜなら、半分の被験者が完全な知識を持っている質問項目について、他の半分の被験者は、その質問項目について全員が、間違った方の選択肢を正しいと思い込んでいるということだからである。

そこで、実際に生じ得る予測・実測差は、(ここで最も極端な場合を仮定したように、被験者が極端に異なる2集団から構成されていたとしても)これより少ない。そして、実際には、そのように極端な集団構造ではなく、各被験者の知識レベルは適度にばらついているはずなので、さらに予測・実測差は小さいと考えられる。実際にSaito (1998)のセッション1において、48項目全体で最低の正答率(58%)を示した被験者は49名中5名、90%以上の正答率を示したのは2名であり、あとは全体の平均正答率(73%)の周りにばらついて

いる。

問題は、このばらつきが項目ごとに全く異なる傾向を示しているかどうかである。言い換えると、 a の値が質問項目ごとに大きく異なっているならば、信号検出理論から導かれる式(1)は全く当てはまらず、Figures 3や4のような結果になるであろう。残念ながら、現有データでは個々の項目ごとに被験者の知識レベル(信号検出理論のstrength)を推定できない。各被験者は同一項目に対して2回しか解答していないため、この推定を行うには少なすぎる。しかも、推定が可能な程度のデータを得るために何度も(同一質問セットに)解答させると、しだいに被験者は自分の(以前のセッションでの)解答についてエピソード記憶を形成してしまい、分布の推定自体ができなくなると考えられる。このため、(少なくとも現段階では)理論的に何がもっともらしいかを議論し、暫定的な結論を導くしかない。

さて、Saito (1998)で用いられた質問項目は、ヨーロッパの都市名が呈示され、その都市が属する国名2つが選択肢とされた。このとき、選択肢となる2つの国名はイギリス、ドイツ、フランスの中から2つ、また、イタリア、ギリシャ、スペインの中から2つのように、どちらかの選択肢セットから抽出されていた。48の都市名は、各国について8個ずつであり、比較的日本人が知っている可能性のあるものから、そうでないものまで出来るだけ均等になるように選ばれていた。つまり、予備的調査によって、正答率が平均75%程度(確率水準である0.5と完全正答の1.0の間)になるように、かつ、できるだけばらつきが大きくなるように選ばれていた(知覚尺度を構成するための実験と同様に、項目間の相違があつて確信度が低い値から高い値まで出るようになっている)。Figure 3や4をみれば、各項目の正答率は適度にばらついており、かつ特に偏りはないのが分かる。

さらに、被験者はすべて特定大学において児童学を専攻している3年生の女性である。特定の大学へ進学した、特定の学科の、殆ど同年齢の同性集団だということになる。被験者の集団としては非常に均質である

と考えられよう。このようになってくる場合、項目によって特別に、被験者が極端に異なる集団に分離してしまうとは考えにくい。例えば、被験者の中にヨーロッパ系の外国語学科の学生あるいはそのあたりに関する地理学を専門にしている学生が含まれていたなら、スペイン語やイタリア語などの発音の特徴から（かなり高い確率で正しい）解答を推測するであろう。このような被験者が（例えば半数）含まれていると、被験者集団の中に、極端に異なるタイプのサブグループが出来てしまい、上記のような大きな予測・実測差が出ることも考えられよう。しかし、そのような場合でも、項目ごとにFigures 3や4のように予測・実測差が、項目間でランダムになるとは考えにくい。そのように特定地域の地理や言語に詳しい人が、一定割合で含まれていた場合、（この実験で用いられたような特定分野のみに関わる知識の）質問セットでは、詳しい人とそうでない人の割合が、項目ごとに大きく異なってしまふ可能性は小さい。そして、予測・実測差がランダムになるには、ある項目に関する知識について、被験者が極端に異なるサブグループに分かれてしまった項目と、そうでない項目が（その分離の程度が）大きく異なるときに限られるので、この可能性については、小さいと判断するのが妥当であろう。

この実験のように、これほど均質に近い被験者集団であった場合、どの項目でも、よく知っている人から、そうでない人までほぼ均等に分布していた可能性の方が大きい。項目によって、知識レベルの相違は当然あるとしても、被験者の集団としてみると、どの項目にも、様々な知識レベルの被験者が、同じように分布していたと考えられる。つまり、ある項目について（他の人は殆ど知らないのに）ある人だけが特別よく知っているものがあつたとしても、その項目についてだけ特別に集団がサブグループに分かれてしまうということではなく、他の項目については、別の人がよく知っているというように、統計的にならされてしまう（特に集団に構造はない）と考えるのが自然である。

ただし、このように被験者集団内のサブ構造がなかったとしても、被験者間の相違はある以上、式（1）

の予測値に対して、実測値は小さくなる傾向を示すはずである。しかし、上述のように、被験者間の相違は統計的に（項目間で）ならされてしまうので、いくつかの項目のときだけ、はっきりした集団構造に偶然なってしまう確率は小さい。そのようなことが（かなりの数の項目で）起こらなければ、予測・実測差は（ある程度のばらつきはあつても、ほぼ）系統的になる。つまり、Figures 3や4とは異なり、実測値は予測値の下に曲線（に近いパターン）を描くはずである。また、集団内のサブ構造があつても、それが項目間で一貫していたなら、同様のことが起こる。要するに、集団にサブ構造があり、かつ、その集団構造が項目間で（かなり大きく）異なっているときにのみ、Figures 3や4のような結果が出ると考えられる。しかし、上記の考察に基づけば、すべての質問に同じ被験者が答えているので、そのようなことが起こっていた可能性は小さいと判断してよいであろう。

結論と要約

一般知識問題を用いた2肢選択課題の実験について、知識の検索結果から選択肢を選ぶ過程の理論として、これまで信号検出理論が基本的枠組みとして用いられてきた。実際には、個々の知識理論によって、具体的な理論の内容は異なるのであるが、その基本的な部分に信号検出理論の枠組みが用いられている。この枠組みそのものを検証する試みとして、正答率と解答変更率の関係が予測され、Saito（1998）のデータを質問項目別に集計しなおして、予測と実測の一致が検討された。

その結果、予測を否定する結果が得られたが、そのようになった理由として、被験者間の知識レベルの相違が指摘され、これによって生じる予測・実測差が推定された。最も極端な場合、解答変更率がゼロになってしまう程度まで、予測・実測差が出ることもあり得ることが証明されたが、しかし、そのようなことが起こる可能性は殆どない（実際に実測の解答変更率はゼロではない）。被験者の知識レベルに相違があれば、実測は予測より小さくなるのは間違いない。しかし、

問題は、この予測・実測差に法則性がまったく見られないことである。

このような結果が生じる原因について検討されたが、ある条件が満たされたときにそうなるという結論であり、その条件（集団内のサブ構造と、その構造の項目間の大きな相違）が、実際に満たされていたとは考えにくいという議論がなされた。もし、この結論が正しければ、記憶理論（少なくとも意味記憶・知識理論）における、判断系の部分は、大きな問題を抱えることになってしまう。

しかし、本稿ではそこまでの強い結論は控えたい。確率変動を伴うシステムや集団の振る舞いは、詳しく検討すると、思いがけない性質が発見されることも珍しくない（e.g., Townsend & Ashby, 1983）。例えば、信号検出理論の枠組みが基本的には採用されていても、Ratcliff（1981; 1985）のような出力系にランダムウォークの性質を持っているような理論であった場合、かなり複雑な個人差が（条件によっては）予測されるかもしれない。この問題に関する最終結論までには、かなりの理論的研究が必要になると思われる。

そこで本稿としては、少なくともある特定の実験パラダイムにおいて、信号検出理論の基本形に基づく予測と合致しない現象が発見されたこと、ただし、確率現象の組み合わせでそれを説明できる可能性がないとは言えないという結論になる。

References

- Barakrishnam, J.D., & Ratcliff, R. 1996 Testing models of decision making using confidence ratings in classification. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 615-633.
- Bernback, H.A. 1967 Decision processes in memory. *Psychological Review*, 74, 462-480.
- Donaldson, W., & Murdock, B.B.Jr 1968 Criterion change in continuous recognition memory. *Journal of Experimental Psychology*, 76, 325-330.
- Glanzer, M., Adams, J.K., Iverson, G.J., & Kim, K. 1993 The regularities of recognition memory. *Psychological Review*,
- Kintsch, W. 1967 Memory and decision aspects of recognition learning. *Psychological Review*, 74, 496-504.
- Murdock, B.B.Jr, & Dufty, P.O. 1972 Strength theory and recognition memory. *Journal of Experimental Psychology*, 94, 284-290.
- Norman, D.A., & Wickelgren, W.A. 1965 Short-term recognition memory for single digits and pairs of digits. *Journal of Experimental Psychology*, 70, 479-489.
- Murdock, B.B.Jr. 1974 Human memory: Theory and data. Hillsdale, N.j.: Erlbaum Press.
- Ratcliff, R. 1981 A theory of order relations in perceptual matching. *Psychological Review*, 88, 552-572.
- Ratcliff, R. 1985 Theoretical interpretation of the speed and accuracy of positive and negative responses. *Psychological Review*, 92, 212-215.
- Saito, M. 1998 Fluctuations of answer and confidence rating in a general knowledge problem task: Is confidence rating a result of direct memory-relevant-output monitoring.? *Japanese Psychological Research*, 40, 92-103
- 妻藤真彦 1999 確信度評定に関する標準仮説と対立仮説の理論的検討. 美作女子大学・美作女子大学短期大学部紀要, 44, 8-17.
- 妻藤真彦 2000 確信度理論の展望 美作女子大学・美作女子大学短期大学部紀要, 45, 1-9
- Swets, J.A., Tanner, W.P., Jr, & Birdsall, T.G. 1961 Decision processes in perception. *Psychological Review*, 68, 301-340.
- Tanner, W.P., Jr, & Swets, J.A. 1954 A decision-making theory of visual detection. *Psychological Review*, 61, 401-409.
- Townsend, T.S., & Ashby, F.G. 1983 *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- Wallsten, T.S. & Gonzalez-Vallejo, C. 1994 Statement verification: A stochastic model of judgment and response. *Psychological Review*, 101, 490-504.

(2000年12月1日 受理)