

**確信度評定過程の量的検討：  
正答率から予測される確信度出現頻度分布**

**妻 藤 真 彦**

美作大学・美作大学短期大学部紀要（通巻第49号抜刷）

論 文

## 確信度評定過程の量的検討：正答率から予測される確信度出現頻度分布 Confidence-rating mechanism: Frequencies of confidence values predicted from each correct answer rate.

妻 藤 真 彦

確信度評定のメカニズムは、認知心理学の中の記憶研究、特に再認記憶の分析では、信号検出理論の枠組みの中で理論化されてきた。もともと感覚系の刺激閾概念への批判に基づき、その代替として提案された信号検出理論 (Signal Detection Theory: e.g., Tanner & Swets, 1954; Swets, Tanner, & Birdsall, 1961) は、心理学実験の被験者が何らかの意思決定・判断を行うときの基本理論として広く応用されてきた (e.g., Bernback, 1967; Donaldson & Murdock, 1968; Kintsch, 1967; Murdock & Dufy, 1972; Norman & Wickelgren, 1974, for review)。そして特に記憶研究の分野では、確信度の評定メカニズムについて、70年代にはすでに研究対象というよりも、この理論に基づく確信度データ解釈が研究法の一部だとされた (e.g., Murdock, 1974)。ここでは、確信に関する哲学に近い議論 (妻藤, 1992) は避け、この記憶に関する確信の機能的側面 (確信評定のメカニズム) について検討する。

再認記憶の研究では、記憶系の出力が何らかの一次元の量 (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大きければ「見たことがある」そうでなければ「ない」と判定されるというように、基本的に信号検出理論を用いてデータが説明され、また解釈されてきた。そして、その「量」と、ある基準値との差に尺度値をあてはめたものが確信度だと「仮定」される (以下では、これを標準仮説と呼ぶ)。これらは時にはそのことに言及することもなく「使われ」ている (e.g., Barakrishnam & Ratcliff, 1996; Glanzer, Adams, Iverson, & Kim, 1993; Wallsten & Gonzalez-

Vallejo, 1994; 妻藤, 2000, for review)。

しかし, Saito (1998) は, 2肢選択による一般知識問題を2回繰り返した時の確信度の確率変動パターンが確信度評定の標準仮説による予測と正反対のものになることを示し, 代替案として「ふらつき仮説」を提案した (また妻藤, 1999b, はこの現象の頑健性を確認している)。この「ふらつき仮説」は結果のパターンについて説明できるものであるが (妻藤, 1999a), 量的な予測についてはまだ検討されていない。

質的な説明のみであれば考慮されるべき要因が確信度評定過程のみでよいのに対し, 量的な予測となると関与する可能性のある要因がいくつかあり, それらに関する仮定を入れる必要がある。そのとき, 解答 (再認) そのものに関する判断過程とのかかわりをどのように扱うかも量的検証の要点となる。標準仮説はこの判断過程と切り離すことができないため, 2つの確信度説を量的予測によって比較するには, 判断過程の理論を無視できない。しかし判断過程の理論の基礎になってきた信号検出理論に疑問点もあり (妻藤, 2000; 2003), その上, 個々の記憶系出力の確率変動について一般的に成立する理論もない。

妻藤 (2003) は解答のセッション間変動と確信度および正答率の関係を検討したが, ここでも確信度の理論がどのようなものになるかによって, 判断過程の理論として信号検出理論が妥当であるかどうかの量的な検討はできなかった。つまりこれを否定するだけの根拠は成立しなかった。しかし信号検出理論そのものに対する疑いは別の観点からも提出されており (Bar-

krishnam, 1999), このように確信度評定メカニズムと解答判断過程そのものについて、互いに理論的な問題がからみあっており、量的な予測との一致度を使って、ストレートに2つの仮説を比較対照することが難しくなっている。

しかし、Saito (1998) は、標準仮説では「質的」説明がまったくできない(当然量的予測もできない)データパターンを発見しているのであるから、ここでは量的予測について2つの説を直接比較する必要はかならずしもないと考え、本稿では、解答判断過程を無視しても検討可能な範囲で、確信度評定の「ふらつき仮説」をモデル化し、このモデルを使って基本部分の検証を行う。

### 理論検証に関わる検討要因

この仮説では、記憶判断の過程そのものとは無関係に(それがどのようなメカニズムであれ)、記憶出力が何らかの形でサンプリングされて、その変動の程度に基づいて確信度の評定が行われるというものである。基本的部分はこれだけであり、Saito (1998) のデータの質的パターンはこれで説明することができる。しかし、この仮説の量的側面を検証するには、いくつかの問題(不確定な要因)が入ってくることになる。

まず(1) 確信度判定のためのサンプリング時に、記憶出力がどのような確率分布になるかという問題がある。前述のように、これをデータから推定しようとしても、そのために判断過程の理論が必要になる。しかもセッション間の解答変化率には判断・意思決定の慣性(最初の判断が繰り返し判断の変動を少なくする)があり(Saito, 2003)、(内的過程における)サンプリング変動率は(データに表れる)解答変化率よりも大きいと考えられる。信号検出理論が正しければ正答率から解答変動率が予測できるはずであるが、妻藤(2003)が示したように、この予測は実際のデータと大きく外れており、これはまた別の理論的問題とリンクしてしまう。つまり、ふらつき仮説と標準仮説のどちらがよりよいかを直接検証しようとしても、判断過

程の問題があるために現時点では不可能である。

しかし、ふらつき仮説自体は(標準仮説とは違って)判断過程とは独立であるため、この仮説が量的にデータと一致するかどうかという検討なら行うことができる。つまり、ふらつき仮説では記憶出力の分布の型は関係なく、各選択肢の評定サンプリング時の出現率のみで、確信度の頻度分布を予測できる。しかも2肢選択課題であれば、出現率2つといっても実際の自由度は1である。このように標準仮説とは違って、記憶・判断過程に関する理論を必要としないのである。ただし、この出現率が実測正答率とイコールであるかどうかは、後述のように若干の議論が必要である。

(2) サンプリングが何回行われるのかは、まったくのフリーパラメータである。しかも、これに関して後述のようにかなりの個人差があると考えられる。

(3) 標準仮説にせよ、ふらつき仮説にせよ、どのように確信度評定値が割り当てられるかは、それを示唆する理論すら存在しない。しかも、各確信度評定値間の距離が等しくないことを示唆する研究すらある(実験条件によって、ファンインあるいはファンアウトが生じて、評定値の間の距離が変化する; Stretch & Wixted, 1999)。それだけではなく、評定尺度の両端あるいは中央は特別な意味を持っている可能性もある。後述のように、実際のデータの度数分布を見ると、明らかに両端(確信ありvsなし)に使用頻度の偏りが見られる(このような評定尺度の場合には、人格検査でよく見られる中央反応傾向はむしろ少ないであろう。確信が持てないという答えは、尺度値の中央ではなく下限になっているからである。この点で、両端への反応傾向が強くても、これはいわゆる極端反応傾向ではない)。

以上の点を整理して検討するために、ふらつき仮説に不確定要因の仮定を入れて数値的予測を行う。

### もっとも単純なモデル化

**基本仮定** ふらつき仮説のもっとも単純なモデルは、記憶系出力をある回数サンプリングして、そのときの解答変動率を確信度だとするものである。これは

単純すぎるかもしれないモデルではあるが、これを検討することで少なくとも問題点を整理することができるであろう。また、これで十分なデータとの一致がみられるようであれば、非常に有力な仮説と考えてよいであろう。

まず、上記(1)の問題は直接解決することはできないので、2肢選択における正答率とその1.0に対する補数が各選択肢のサンプリング時出力率に等しいものとする。そしてこれらの比を変動率とする。この仮定は、妻藤(2001;2003)の結果から見て、おそらく厳密には正しくない。これらの分析から見ると、確信度の評定メカニズムについて数値的予測ができるモデルがあれば、ある質問に対する解答として「意思決定」されたときの反応率(正答率)から、(その意思決定とは別に生じる)生の記憶系出力の出現率(およびそこから導かれる出力項目の変動率)を関係づけることができる。しかしそのような確信度評定モデルを検証するには、その数値的予測があらかじめ必要である。双方が互いに相手を必要としているため、そのままでは解決する方法はないことになる。そこで、とりあえずいささか強引であるにも拘わらず、変動率についてはデータから直接計算できるという仮定を近似として使うことによって、確信度評定側の性質を検討するという方法をとらざるを得ない。

以下では、この仮定の上で、(2)と(3)についてモデルの当てはめを行い、その後さらに別のデータセットを予測することで検証する。

### 自由度の問題

ここではSaito(1998)の実験における確信度の使用頻度分布に対する当てはめを行うが、見かけ上データ側の自由度より理論の自由度が大きすぎるように見えてしまうため、若干の議論を行っておく。以下の当てはめでは、48個の質問項目の各々について正答率から推定される確信度の頻度分布を計算する。つまり各質問項目ごとに確信度評定サンプリング誤差の確率を計算して評定値がどのようにばらつくかをシミュレーションする。その48種類の結果を合算して、実験全体

での評定値頻度分布を計算することになる。

最終的なグラフ(48個の質問項目のデータを全て合算した評定値頻度分布)では一見すると5個の独立変数に対する予測値の計算のように見えるが、実際にはそうではなく、予測計算の出発点になる変数は個々の質問項目の正答率である。つまり1個の質問項目について実験で得られた正答率から(以下で述べるように)11個の確率が予測され、それらが質問項目48個分なので576個の予測値が得られる(48個の正答率から576個の頻度が予測される)。この予測値セットについて、同一のカテゴリー(11種類)として集計されるもの同士を48個分合計する(これで11個になる)。この集計に自由度はない。そして同一の評定値になるもの同士も合計して最終的に5個の頻度予測値(確信度の頻度分布)を得る(ここには後述のように自由度がある)。理論側の自由度は見かけとは相当異なっている。

ただし、前述のように現時点では不明なファクターがあり、それらを仮定によって決定しなければならない(同一のカテゴリーになると想定される数値を合計するときに幾つかの仮定が必要になり、また最終的に評定値を割り当てるやり方にも自由度がある)。このため、理論側の自由度がいささか大きいので、1つの当てはめのみでは、(近似モデルであるとはいえ)検証として不十分だと思われる。

そこで、Saito(1998)実験のセッション1のデータに対して当てはめを行い、そこで得られたカテゴリー分割法と評定値割り当て法を使って、セッション2のデータを予測し、そのときの予測とデータの一致度を検証として扱うことにする。

最終的に集計されたセッション1と2のデータは、最終的集計のパターンは似ているが、両セッションの間には解答の変動が20パーセント以上あり、全体の平均正答率(48項目の正答率の平均)に有意差はないにもかかわらず、個々の質問項目での正答率の変動はかなり大きい。そのため、セッション1の各質問項目の正答率を使って予測された確信度評定値の頻度分布は、各項目ごとに見ると大きく違っている。その上、平均確信度のセッション差は小さいとはいえ有意で

あった (Saito, 1998)。そのため同一のカテゴリ・評定値になる (と仮定された) ものを合算した時に、(セッション1での当てはめが良好だったとしても)、セッション2の場合にもデータと一致するとは限らないのである。このような方法をとることで、理論の近似モデル当てはめ結果は検証として十分な意味を持つと考えてよいであろう。

### モデルの当てはめ

モデルの基本部分 ここでは「ふらつき」仮説のもっとも単純なモデルとして、確信度評定の際に記憶のサンプリングが行われ、そのときの記憶出力変動率が主観的確信の程度を決めるものとする。サンプリング回数を  $n$  とし、片方の選択肢が再認された回数を  $k$ 、他方の回数を  $n-k$  とする。この単純なモデルの場合、各変動印象の出現確率は、

$${}^nC_k p^k (1-p)^{n-k} + {}^nC_{n-k} p^{n-k} (1-p)^k \quad (k \neq n-k)$$

$$\text{or } {}^nC_k p^k (1-p)^{n-k} \quad (k = n-k)$$

で決まることになる。 $n$ は定数であり、 $k$ は0から $n$ まで変化する。この異なる $k$ の値 (というより $k$ と $n-k$ の比) が変動印象 (変動率) である。ここでまず、問題になるのは $n$ であるが、これはフリーパラメータとして、データに最も近い確信度値の分布を示す値を求める。

評定カテゴリへの変化率の割り当てとサンプリング回数の推定 非常にやっかいな問題はむしろ (3) であり、この変動印象に対してどのように確信度評定尺度値が割り当てられるかである。実験の指示では7段階尺度で答えるようになっており、確信があるを7、ないが1とされていた。しかし、実際に被験者が7種類の区別までしていたかどうかは別問題である。そこでまず、Saito (1998) のデータから各被験者が2つのセッション中に使用した確信度評定値の数 (種類) が算出された。6種類と7種類使用した被験者がもっとも多いという結果になった (Figure 1) が、実際に被験者が区別していた変動印象は、おそらくこれより少ない。なぜなら尺度の両端は別として中間の値であれば、尺度値割り当てに他の要因 (例えば、7段階の

全てを使用しないといけないように思うなど) がからむことによって、他の数値も使おうとするかもしれない。そこで、1回あるいは2回しか使用しなかった尺度値は、そのようにして生じた可能性が相当大きいと考えられるので、各被験者が2回以上使用した尺度値数の分布を出す (Figure 2) のようになり、最頻値は6個、3回以上使用したものと (Figure 3) のようになり、最頻値は5になる。これら3つのどれが最も適切であるか決定する方法はない。しかし、ここでは、上述のように各項目の出力変動率計算自体を大まかな仮定で近似するしかなく (しかもこの大まかな仮定は各個人ごとに決定することができず、集団データを個人のように扱うしかない) ので、以下の方針をとることにする。

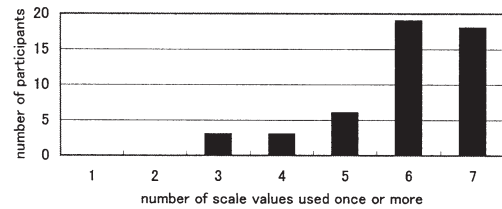


Figure 1. Number of scale values used once or more by each participant in two sessions

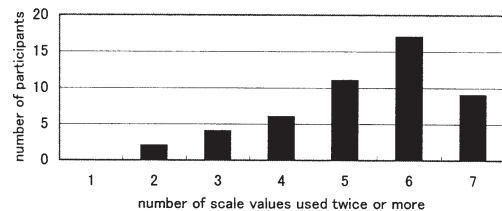


Figure 2. Number of scale values used twice or more by each participant in two sessions

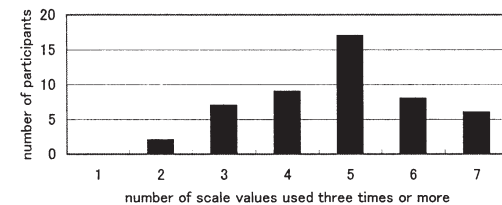


Figure 3. Number of scale values used three times or more by each participant in two sessions

まず、データについては、尺度の両端がアンカーポイントとして重要であるのははっきりしているので、

尺度値1と2の頻度はまとめてしまい、同様に6と7もまとめてしまうことで、5段階尺度であったものとして扱う。実際6と2の使用頻度は他と比べて少なく、被験者によって2や6とせずに1か7にしてしまったケースがかなりあった可能性が強い。

ただし、データ自体をそのようにまとめるのは単に集計の仕方の問題であるが、モデルの側ではそう単純ではない。もちろんデータが5段階にまとめられているとき、モデルの側も尺度値の使用個数(使用種類数)5・6・7をまとめてすべて5段階として予測値を合算するのはデータの場合と同じ意味である。しかし問題は4段階以下の尺度値を用いていた被験者に関しては、別に計算しないと予測頻度に歪みを生じてしまう。例えば、7段階を使用した被験者集団の予測値について6と7をまとめても、データ側も6と7がまとめられておれば、(検証の自由度は下がるが)各々の合計値は(モデルが妥当なら)一致するはずである。しかし、例えば3段階しか使っていなかった場合は、その被験者(群)が3段階について、もし1・3・7という尺度値を割り当てていたなら、全体データの1&2、3、6&7の集計頻度に寄与する一方、この被験者群は4と5という尺度値を使わないために、その分データ全体での尺度値の使用頻度は少なくなるのである。したがって、予測を5段階にまとめたとき、段階数の多い被験者群については5段階で計算すればよいが、4段階以下のケースについては別に計算して合計しなければならない。このときの各段階が含まれる割合は、上記の使用尺度値数の統計に基づいて決定する。例えば4段階使用した被験者の人数比を4段階予測の各頻度に掛けたものを、合計することになる。ただし、上記のFigures 1, 2, あるいは3のどれが妥当であるかという問題が残るが、ここでは、一回しか使っていない場合は偶然とみなし、Figure 2を採用する(ただし、Figure 1による計算も行って比較する)。

もう1つの問題として、モデルにおける出現率にどのように5段階(4および3段階)を当てはめるかがある。これもデータを見ると、全ての項目を平均した正答率が75パーセント前後であるのに対して、確信度

は最大値と最小値の頻度が非常に大きい。ということは、明らかに尺度値の両端にかたよる傾向があると考えてよい。n回のサンプリングが行われるとしたとき、理論上区別できる出現率の種類はnの半分であるが、中程度の変動率はそのまま(6段階の)ミドルレンジ4段階に当てはめ、両端のいくつかの変動率をまとめて、確信度最大あるいは最小に割り当てよう。このとき、データでは確信度最小よりも最大の方が多いので、最大の方により多く割りあてることにする。この点がモデル側の自由度を引き上げてしまうが、しかし明らかに全ての組み合わせの中からデータフィット最大になるように選ぶことは出来ない。つまり、最小の評定値頻度予測を得るために最小と中間の出現率をまとめるような組み合わせは、理論上ありえない。ということは評定のためのサンプリングで決まる出現率の全ての組み合わせの数を、モデルの自由度だと考える必要はない。あくまで、組み合わせが可能なのは、理論上、隣り合う大きさの出現率のみである。

しかし、組み合わせを決定し、それに評定尺度値を割り当てるいかなる規則・法則もなく、またそれを決める理論も存在しないので、ここではまず、上記の両端に関するデータの観察から得られた傾向のみを用いて、暫定的な規則を仮定してしまう。サンプリング回数が何回であれ、最大評定値と最小評定値以外は、1つの変化率を対応させる。例えば20回サンプリングであると、異なる変化率は11種類になる(片方の選択肢のサンプリング出現数と他方の出現数の合計が20になる組み合わせの数)。5段階の尺度とするなら、11から(最大と最小以外の)3を引いた残りの8個の出現数を最大と最小に割り当てる。このとき、最大の方が最小よりも多い傾向があるので、もしこの残りが奇数のときは最大の方が1つ多くなるように割り当てる。5段階の場合は、この残りは偶数なので、最大最小とも4個となる。

まず、サンプリング回数を推定するために、全ての被験者が5段階だったものとして、この規則を適用し、データとモデルの相違が最小になるように(データ分布とモデル分布を比較する $\chi^2$ 乗の値が最小になる)

サンプリング回数決定された。その結果は20であった。

4以下の段階数についても同じ規則で決定すると、次のようになる。4の場合は、最大に上位の4個、中間の2段階は各々2個ずつ、最小は下位3個；3段階では、最大に上位4個、中間段階は4段階のときの中間2個をまとめるので4個、最小は下位3個である。このとき4と3では、中間段階を評定値のどれに割り当てるかを決定しなければならない。1 & 2, 3, 4, 5, および6 & 7の5段階の内、中央の3つのどれかに割り当てることになるが、ここではまず低い評定値の方に割り当てるものとし、セッション1での当てはまりを見る。これが良好であれば、そのままの割り当て法により、セッション2のデータを用いて再検証する。

問題は2段階の場合である。上記の規則によって3段階以上であればひととおりに決まるが、2段階の場合も同じようにしてよいかどうか疑問がある。2段階しかない場合、それらが最大と最小になるとは限らず、中程度を表すものがないのであれば、たとえば2と6のようになる可能性はかなり強いと思われる上、この割り当て方にかなりの個人差があることも考えられる。しかもFigure 2のように、2段階しか使用しなかった可能性のある人数は非常に少ないので、ここでは3段階で近似することにし、3段階を使用したと推定される人数に加算してしまう。

### セッション1の頻度分布

Saito (1998) の素データからセッション1の質問項目ごとの正答率を計算し、48項目の各々について5から3段階の評定値頻度分布予測値を算出した。そして、Figure 2のデータを用い、5段階以上を用いた人数を43名、4段階を6名、3段階以下を6名として全体の頻度分布にしたものがFigure 4に、実測値と共に示されている。横軸の確信度1は、元データの1 & 2, そして5は元データの6 & 7を示す。モデルの予測値と実測値の間に有意差はない ( $\chi^2=6.91$ ,  $df=4$ ,  $p > 0.10$ )、また一致係数は十分に大きい ( $r^2=0.99$ )。

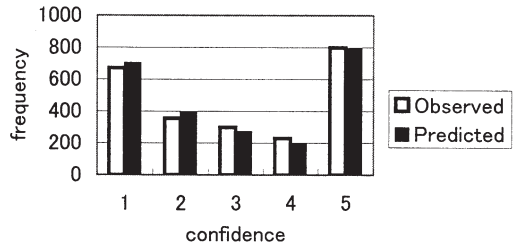


Figure 4 Frequencies of confidence values in session 1.

4段階と3段階の人数についてFigure 2ではなくFigure 1の、一回以上用いた回数によって人数を設定すると、カイ2乗の値はさらに小さくなった ( $\chi^2=4.97$ ,  $df=4$ ,  $p > 0.20$ ;  $r^2=0.99$ )。しかし、前述の考察のように一回しか使用していない評定値は評定値割り当ての揺らぎによって生じただけである可能性が高いこと、またセッション1のデータへの当てはめすぎ(偶然的変動まで含めて当てはめてしまうこと)を防ぐために、セッション2データへの当てはめでも、Figure 2の方を優先する。

いくつかの仮定と評定過程サンプリング回数が設定されたが、それらには、セッション1の当てはめがもう少しうまくいかないときには、変更できる側面がいくつかあった。つまりこの点でモデル側の自由度が大きくなっており、ここでの当てはめだけでは十分な検証とはいえないという批判が考えられる。そこで、ここで当てはめが良好であった設定をそのまま使い、セッション2のデータを用い、そこでも良好なモデルとデータの一致が見られるかどうかを検討する。

### セッション2の頻度分布

セッション1での設定はそのままにして、セッション2の各項目の正答率48個を独立変数としてモデルに投入し、同じ方法で計算が行われた。その結果は実測値とともにFigure 5に示されている。ここでも当てはまりは良好で、一致係数は大きく ( $r^2=0.99$ )、またモデルと実測値の間に有意差はなかった ( $\chi^2=6.61$ ,  $df=4$ ,  $p > 0.10$ )。段階数の人数を一回以上使用した尺度数で決めると (Figure 1)  $\chi^2$ 乗の値はさらに小さ

い ( $\chi^2=4.24, df=4, p>0.30$ ;  $r^2=0.99$ )。

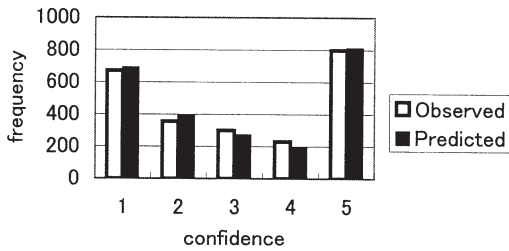


Figure 5. Frequencies of confidence values in session 2.

上述のようにセッション間では20パーセント以上の解答変更があり、平均正答率に有意差はないものの各項目の正答率はセッション間でかなり異なっている。しかも平均確信度のセッション差は有意であった。したがって、セッション1でのデータについて行われたモデル当てはめが、セッション2での確信度データにも当てはまるかどうかは自明ではなく、検証としての意味を十分にもっている。

サンプリング回数20回は、異なるデータセットに対しても有効であった。そして11種類のサンプリング変動率は、サンプリング回数のみをパラメータとして決まる。あとは、変動率の評定尺度へのまとめ方と評定値そのものへの割り振りであるが、これはセッション1で決定されたものがそのまま使われたので、セッション2の正答率セットに対するパラメータ推定は全くなかった。この意味でセッション2のデータとの良い一致が見られたのは、ふらつき仮説のこのモデルが有効な予測をしたと結論してよいであろう。

## 要約と結論

確信度評定過程のふらつき仮説を検証するため、解答判断過程とは独立に検討できる最も単純なモデルが設定され、Saito (1998) の2肢選択一般知識問題における確信度使用頻度分布が当てはめられた。この当てはめは、ひとつの質問項目の正答率から各確信度値の出現確率分布が予測され、これらを48個の項目について合算したものがデータと比較された（このモデルでは、2肢選択課題の場合、正答率がひとつ与えられ

ると、その質問項目に関するすべての確信評定値の出現確率を予測する)。ひとつのデータセット (セッション1) を用いて、パラメータと評定値割り当て規則が決定され、データと予測の良い一致が見られた。ここで決定されたパラメータ等をそのまま使い、同じ被験者群によるもう1つのデータセット (セッション2) に対して、最初のデータセットで決定されたパラメータ等をそのまま使って、ここでも良い一致が得られた。後者のデータセットには、前者に対して平均20パーセント以上の解答変動があり、各質問項目の正答率はそうとう大きく違っている。また平均確信度には有意差があるので、この2つめの予測とデータの一致は、検証としての意義が大きいと考えられる。

ただし、セッション1で特定の確信度であったときの、セッション2での条件付確信度分布については、まだ検討できていない。これを予測するためには、確信度の変動にかかわる別の補助仮説が必要になるからである。Saito (2003) は、確信度変動をある程度抑制するファクターがあることを示す実験結果を出しており、これに関する理論的研究が追加されなければ、これらの条件付分布を計算することができない (すべての確信度に対して同じ強さで抑制ファクターが働くなら単純であるが、それを示す証拠はない)。

しかし、まだ一部検証が必要な部分が残っているとはいえ、ふらつき仮説の最も単純なモデルでもかなり良く確信度使用頻度分布を予測できたこと、さらに標準仮説ではSaito (1999) のデータを質的にもまったく説明できないことから考えて、かなり有望な代替仮説であると結論してよいと思われる。

## 引用文献

- Barakrishnam, J.D. 1999 Decision processes in discrimination: Fundamental misrepresentations of signal detection theory. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1189-1206.
- 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. 1974 *Human memory: Theory and data*. Hillsdale, N.j.: Erlbaum Press.
- 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.
- 妻藤真彦 1992 根拠を述べるできない確信と「意識様態」. 美作女子大学・同短大部紀要, 38, 1-10.
- 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 1999a 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 1999b 一般知識問題における確信度変動分布と理論的問題. 日本心理学会第63回大会発表論文集, p626.
- 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
- 妻藤真彦 2001 再認記憶の理論における信号検出パラダイムの問題点. 美作女子大学・美作女子大学短期大学部紀要, 46, 16-21.
- Saito, M. 2003 Two modes of confidence rating: An effect of episodic information of participant's own past responses in a repeated-question paradigm. *Japanese Psychological Research*, 45, 94-99.
- 妻藤真彦 2003 信号検出理論と確信度：解答の変動と正答率の理論的問題. 美作女子大学・美作女子大学短期大学部紀要, 48, 8-15.
- Stretch, V., & Wixted, J.T. 1999 Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1397-1410.
- 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.
- Wallsten, T.S. & González-Vallejo, C. 1994 Statement verification: A stochastic model of judgment and response. *Psychological Review*, 101, 490-504.