

二者択一予測判断課題における期待正答率が確信度見積に及ぼす効果

松永, 勝也

原口, 敏一
九州大学文学部

<https://doi.org/10.15017/2328537>

出版情報 : 哲學年報. 47, pp.1-14, 1988-02-29. 九州大学文学部
バージョン :
権利関係 :

二者択一予測判断課題における 期待正答率が確信度見積りに及ぼす効果

松永勝也・原口敏一

はじめに

不確実な場面において判断を行うとき、人がその判断に対して抱く確信 (confidence) の度合 (以下、確信度) はさまざまである。人が抱く確信度の妥当性は、数値による確信度の見積とその確信度の見積がなされたすべての判断における実際の正答率とを比較することによって評価することができる (Adams, 1957; Adams & Adams, 1961)。

Lichtenstein & Fischhoff (1977) は、被験者に択一式の判断課題を多数与え、正しいと思う選択肢を選ばせ、その判断に対する確信度を確率で見積らせた。人の確信度見積りが妥当なものであれば、「この判断が正答である確率は p である」と見積られた n 個の判断のうち np 個が正答になると数学的に期待される。しかし、実際には、確率見積りによる確信度とその確信度による判断の正答率 (hit rate) とは一致せず、見積られた確信度が高いときほど確信度が実際の正答率を上回る過剰確信 (overconfidence) となると報告されている (Lichtenstein et al., 1977)。Fischhoff, Slovic, & Lichtenstein (1977) の研究でも同様の結果が得られ、被験者が極端な確信度である 1.00 を見積ったときでさえ 20% もの誤答があった。

高い正答率が期待できる判断課題では、高い確信度が抱かれるであろう。そうであれば、高い正答率が期待される判断課題でほど過剰確信が見られる

表1 期待正答率の計算法

$$.05 \leq p \leq 1.00, 0.00 \leq q \leq 1.00$$

事象	生起確率	予測	予測比率	結果の比率	当否
A	p	A	q	p q	当
		B	1-q	p-pq	否
B	1-p	A	q	q-pq	否
		B	1-q	1-p-q+pq	当

$$\text{期待正答率} ; \text{epc} = 1 - p - q + 2pq$$

と考えられる。そこで、本研究では、判断課題の期待正答率 (epc ; expected proportion correct) が確信度の見積に及ぼす効果を検討することを目的とした。

実験にあたっては、判断課題として2事象のうちいずれが生起するかを予測する試行を繰り返す二者択一予測判断課題を用い、確信度の測度として確率を用いた。期待正答率は、その予測判断課題を構成する2事象の生起確率を変えることによって操作した。

しかし、二者択一予測判断課題では、被験者がどのような比率で2事象を予測するか(予測比率)によって期待正答率は変わってしまう。表1に示すように、期待正答率は事象の生起確率と予測比率の関数だからである。たとえば、サイコロを投げて6以外の目が出るか6の目が出るか(生起確率は.83と.17 ; $p = .83$)を予測する判断課題では、6以外の目と6の目を1:1(予測比率は.50と.50 ; $q = .50$)で予測した場合、期待正答率は.50であるが、3:1(予測比率は.75と.25 ; $q = .75$)で予測した場合では.67である。したがって、被験者がどのような比率で2事象を予測するかを無視して確信度の見積に及ぼす期待正答率の効果を検討することはできない。そこで、本研究では、被験者が常に最頻事象を予測した(予測比率は1.00と0.00 ; $q = 1.00$)と仮定し、被験者が見積った確信度をすべて最頻事象の予測に対する確信度に換算することによって期待正答率を予測判断課題ごとにそろえた。すなわち、被験者が非最頻事象を予測し確信度を p と見積ったときの確信度は、 $1-p$ として処理した。この仮定の下では、期待正答率は最頻事象の生起確率と等

しくなり、.50 から 1.00 まで変化する¹⁾。

実験結果の分析にあたっては、

1. 被験者の予測方略、
2. 試行の進行にともなう確信度の推移、
3. 期待正答率の変化にともなう確信度の変化、

の3つを検討した。

方 法

〈被験者〉

大学学部生 5 名 (男子 2 名, 女子 3 名: 年齢 21 歳~23 歳) が被験者であった。彼らはランダムな順で 4 つの実験条件に 1 回ずつ参加した。

〈装 置〉

図 1 に示すように、マイクロコンピュータ (MZ-2200 ; SHARP) に接続さ

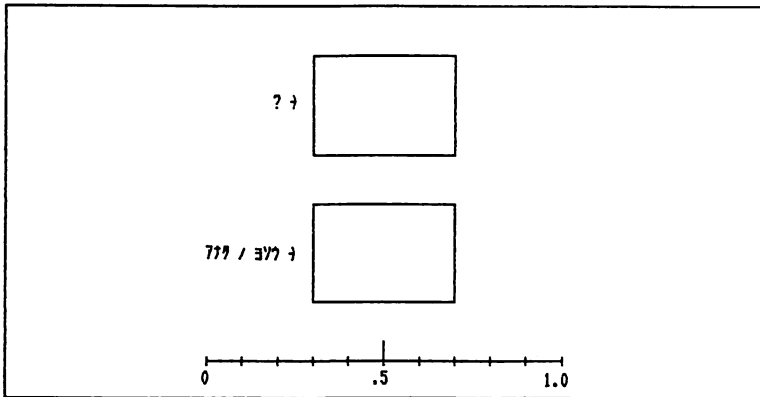


図 1 被験者に提示した CRT 画面の配置

1) 表 1 の期待正答率の計算式; $epc = 1 - p - q + 2pq$ において, $q=1$ とおけば, $epc = 1 - p - 1 + 2p = p$ 。したがって, 常に最頻事象を予測すれば, 期待正答率は最頻事象の生起確率と一致する。

れている CRT 画面上に 2 個の白色矩形を縦に並べて提示した。被験者は、キー操作によって下の矩形を赤色あるいは青色に変えることができた。下の矩形の下に 0.0 から 1.0 まで .1 ステップで目盛った数直線を提示した。被験者は、キー操作によって数直線に直交する縦棒を左右にスライドさせて確信度を表示することができた。被験者に確信度の表示を求めるとき、数直線に直交する縦棒の初期値は .5 の位置に設定されていた。

〈実験条件〉

本実験における実験条件は、予測判断課題の期待正答率である。期待正答率を操作するために、上の矩形に提示する赤色と青色の生起確率を（赤色，青色）= (.90, .10), (.80, .20), (.70, .30), (.60, .40), となるようにした。すなわち、期待正答率の条件は、.90 条件、.80 条件、.70 条件、.60 条件、の 4 つであった。

各予測判断課題の試行数は 300 試行であった。あらかじめ作製しておいた試行系列にしたがって赤色あるいは青色のいずれかを順次提示した。試行系列は次のようにして作成した。実験条件を満たす数の赤色と青色で 30 試行分のブロックを作り、これをマイクロコンピュータ内蔵の疑似乱数を使ってランダムな配列にした。同様の手続きで多数のブロックを作り、そのなかで相互に類似していないと思われるブロックを 10 個連結して 1 つの試行系列とした。

〈手続き〉

被験者を CRT 画面の前に座らせ、「上の矩形にあらかじめ定めておいた順番で赤色あるいは青色のいずれかが提示されます。提示される色を予測し、できるだけたくさん当てて下さい」と教示した。キー操作を説明しながら、赤色と青色が交互に提示される練習試行を 20 試行おこない、本試行に入った。

1 試行の構成は次のとおりであった。まず、CRT 画面上に白色矩形が提示された。被験者は、予測の入力を促すアラーム音が与えられた後、予測する赤色あるいは青色のいずれかをコンピュータのキーボードより入力し予測す

る色を表示した。その後、試行系列にしたがって赤色あるいは青色のいずれかが上の矩形に提示された。予測が当たっていれば「当たり」、違っていれば「ハズレ」と被験者は口頭で結果を報告した。その後、再び画面を白色矩形に戻し、次の試行に移った。

各ブロックの最終試行では、被験者が予測する色を表示した後、この予測に対する確信度を .5 から 1.0 の範囲で数直線上に表示するように求めた。このとき、被験者に対し次のような説明を行った。この数直線は今の予測が当たっている確率を見積るためのもので、.5以下は予測しなかった色の方が当たっている確率が高いことになるからありえない。数直線の 1.0 は確実に当たっていることを、.5 は全く見当がつかず当て推量であることを意味する。

実験の進行および反応の記録は、マイクロコンピュータによって行った。被験者に予測の当否を報告させたのは、当該の予測判断課題に集中させるためであった。また、被験者は、数直線が意味することを容易に理解した。

結果と考察

本研究では、(1) 被験者が予測した最頻事象（赤色）の相対頻度（反応確率）、(2) 被験者が実際に経験した正答判断の相対頻度（実際の正答率）、(3) 被験者が見積った確信度、について分析・検討を行った。実験に使用した試行系列は、ブロック（30 試行）ごとに実験条件が保証されていたので、反応確率および実際の正答率についてはブロック単位に集計したデータを、確信度についてはブロックの最終試行で被験者が見積った確率を分析資料とした。

1. 被験者の予測方略について

本実験で設定した二者択一予測判断課題は、いわゆる確率学習事態である。確率学習事態では、常に最頻事象を予測する最大化方略 (maximizing strategy) が採られるとき、期待正答率は最頻事象の生起確率と等しくなり、最大となる (Peterson & Ulehla, 1965)。本研究では、この方略が採られたと仮定して各予測判断課題の期待正答率を操作したが、実際にこの方略が採られ

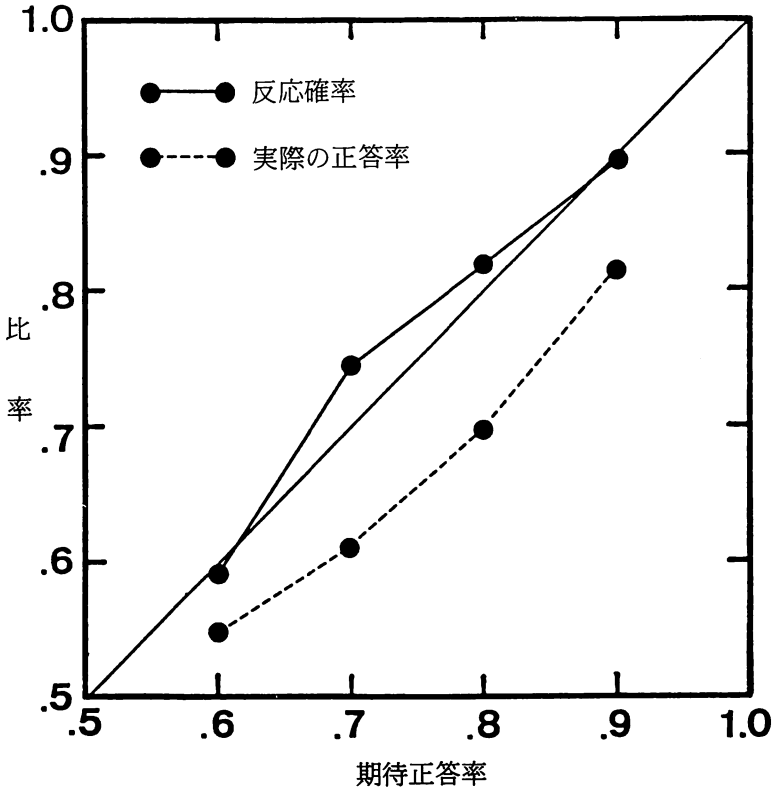


図2 期待正答率の関数としての反応確率と実際の正答率(傾き45度の直線是最頻事象の生起確率を示す)

ることは希である (Rose & Vitz, 1966; Beach & Swenson, 1967)。そこで、被験者が各予測判断課題でどのような予測方略を採用したかを検討した。

図2に、期待正答率の関数としての反応確率および実際の正答率を示す。図中の傾き45度の直線は、最頻事象の生起確率が期待正答率と一致していることを示している。図から分かるように、期待正答率の増加にともなって反応確率、実際の正答率ともに増加している。また、実際の正答率は常に期待正答率を下回っているものの、いずれの予測判断課題においても過半数の予測が正答であったことを示している。

反応確率について一元配置の分散分析を行ったところ、期待正答率の効果が認められた ($F_{(3,12)}=17.836, P<.01$)。さらに、下位検定を行ったところ、5%水準ですべての条件の組合せに有意な差が認められた。このことから、被験者はそれぞれの予測判断課題の期待正答率の違いを鋭く識別し、各予測判断課題の条件に応じて異なる予測方略を用いたと言える。さらに、反応確率は最頻事象の生起確率と極めてよく一致しており ($\chi^2=.107, df=3, .990<P<.995$)、期待正答率と反応確率との間には強い直線関係が認められた ($Y=.994800X+.161999, r=.981, df=2, P<.05$)。したがって、被験者は、当該の予測判断課題を構成する2事象の生起確率の比率と同じ比率で予測するという方略を採用したと言える。

この方略が採られるとき、実際の正答率は最頻事象の生起確率の2次関数となり、(.50, .50)と(1.0, 1.0)の2点を通る下に凸の2次曲線 $2p^2-2p+1$ となることが期待される²⁾。実際の正答率がこの2次曲線に非常によく近似した ($\chi^2=.116, df=3, .980<P<.990$) ことから、被験者が生起確率の比率と同じ比率で予測したことが支持される。

実際の正答率についても一元配置の分散分析を行ったところ、期待正答率の効果が認められた ($F_{(3,12)}=62.402, P<.01$)。下位検定でも、5%水準ですべての条件組合せで有意な差があった。このことは、期待正答率の操作が充分に実際の正答率を制御し得たことを示すものである。しかし、この結果から、実際の正答率が確信度の見積りに及ぼす効果を直接的に論じることはできない。確信度の見積への影響が、実際の正答率によるものなのか反応確率によるものなのか識別できないからである。そこで、既述のような反応確率をそろえる工夫が必要だったのであるが、本研究の確信度に関する結論は実際の正答率のみを独立変数とする研究でも支持されるであろうと思われる。

2) 表1の期待正答率の計算式； $epc=1-p-q+2pq$ において $q=p$ とおけば、 $epc=1-p-p+2p^2=2p^2-2p+1$ 。したがって、2事象の生起確率と同じ比率で予測をすれば、期待正答率は最頻事象の生起確率の2次関数となる。

2. 試行の進行にともなう確信度の推移について

いわゆる確率学習事態では、最頻事象の反応確率は試行の進行にともなう最頻事象の生起確率に漸近するかそれを僅かに上回ると報告されている (Edwards, 1961)。この現象は一般に確率対応現象 (probability matching phenomenon) として知られている。本実験で反応確率が最頻事象の生起確率と一致したことは、この確率対応現象で説明できよう。さらに、Neimark & Shuford (1959) は、最頻事象の生起確率の見積も試行の進行に伴って最頻事象の生起確率に漸近し、確率対応現象を起こすことを報告した。また、Beach, Rose, Sayeki, Wise, & Carter (1970) は、反応確率よりも最頻事象の生起確率の見積の方が早く最頻事象の生起確率に漸近するという結果を得た。以上のことから、被験者が認知した最頻事象の生起確率が試行系列の初期において変動し確信度の見積に影響を及ぼすことが考えられる。

また、Adams & Adams (1958), Koriat, Lichtenstein, & Fischhoff (1980) によれば、確信度とその確信度が見積られた判断の正答率との不一致は、被験者にどの程度不当な確信度が見積られたかを具体的にフィードバックしたり、その確信度を見積った根拠や反例を答えさせたりすることによって修正が可能で、その不一致を小さくすることができる。本実験では、このような手続きを用いなかったが、予測が正答であったか否かは試行ごとに被験者にフィードバックした。予測の当否のフィードバックが確信度を修正する効果を持つならば、確信度は期待正答率に漸近すると考えられる。そこで、試行の進行に伴う確信度の推移について検討した。

表 2 に、各予測判断課題における確信度の推移をブロック順に示す。予測

表 2 条件ごとのブロック (30試行) の最終試行における確信度

期待正答率	ブロック番号									
	1	2	3	4	5	6	7	8	9	10
.60	.58	.69	.61	.64	.65	.65	.68	.65	.66	.70
.70	.77	.61	.76	.74	.77	.74	.69	.79	.70	.70
.80	.77	.72	.81	.68	.82	.72	.87	.82	.83	.85
.90	.77	.76	.74	.80	.72	.69	.71	.82	.73	.74

判断課題ごとに一元配置の分散分析を行ったところ、いずれの予測判断課題においてもブロック間に有意な差は認められなかった。このことから、第 30 試行から第 300 試行の間で確信度は有意な変化をしないと考えられよう。しかし、これは確信度の変化を完全に否定するものではない。

各予測判断課題の第 1 試行では、被験者は予測判断を行うに足る経験は何もしていない。したがって、このときの予測は当て推量で確信度は .50 であったはずである。それが、第 30 試行には .50 より大きな確信度が見積られ以降変化を認めない。このことから、第 1 試行から第 29 試行までの間に確信度の変化過程、恐らくは単調な上昇過程が存在していたと考えられる。しかし、本実験の結果からは、この過程を論じることはできない。

また、表 2 の確信度は、いずれの予測判断課題においても期待正答率からの逸脱がかなり見られる。確信度の見積りが予測の当否のフィードバックによって修正されなかったと言えよう。被験者は、第 1 試行から第 29 試行までの間に形成した確信度を第 30 試行目以降も維持すると考えられる。この点から、試行系列初期の確信度形成過程は重要な意味を持っており、今後検討されなければならない課題と言えよう。

3. 期待正答率の変化にともなう確信度の変化について

既に示したように、期待正答率が高くなるにつれて、最頻事象を予測する相対頻度（反応確率）は直線的に増加し、実際の正答率も 2 次関数的に増加する。素朴に考えて、高い正答率が期待され実際に高い正答率が経験される予測判断課題でほど、被験者は最頻事象の予測に対し高い確信度を見積ると思われる。そうであれば、確信度は期待正答率の増加に伴って単調に増加すると考えられる。また、本実験では、被験者が常に最頻事象を予測すると仮定して被験者が見積る確信度はすべて最頻事象の予測に対する確信度に換算した。したがって、被験者が見積る確信度が正確で妥当なものであれば、確信度は期待正答率と一致し、両者の間に直線的関係が認められるはずである。そこで、期待正答率の変化にともなう確信度の変化について検討した。

図 3 に、期待正答率と確信度の関係を示す。図中の傾き 45 度の直線は、理

論上の確信度の推移を示している。図から分かるように、確信度は期待正答率の増加にともなって直線的に増加してはいない。むしろ .90 条件では .80 条件と比べて減少している。一元配置の分散分析を行ったところ、期待正答率の効果が認められた ($F_{(3,12)}=9.129, P<.01$)。さらに、下位検定を行ったところ、.60 条件-.70 条件、.60 条件-.80 条件、.60 条件-.90 条件、の間に 5%水準で有意な差が認められた。この結果は、期待正答率が高い予測判断課題では確信度が抑制されることを示している。.70 条件-.80 条件、.70 条件-.90 条件、.80 条件-.90 条件、の間に有意差が認められなかったことから、この抑制効果は期待正答率が高い課題でほど強く作用すると考えられる。

確信度を抑制する要因として、「小数の法則 (the law of small numbers)」が考えられる。この法則は、古くから「賭博者の錯誤 (gambler's fallacy)」として知られているもので、偶然性に支配される事態で、ある事象が連続すると次は他の事象が生起するに違いないと考える傾向が人にはある (Tversky & Kahneman, 1971)。本実験の場合、期待正答率が高い予測判断課題のときほど、被験者は試行系列の全体に渡って最頻事象が連続して生起するのを経験した。このため、次に非最頻事象が生起するのではないかという「疑い」を慢性的に抱き、最頻事象の予測に対する確信度が抑制されたと考えられる。一方、期待正答率が低い課題では、2 事象が適当に混合して生起するため「疑い」が抱かれ難かったと考えられよう。

不確定な場面において、非常にありそうな事象を予測するとき、人は予測しなかった他方の事象を強く意識し、当該の予測に対しかえて「疑い」を抱く。そのため、確信度が抑制されると考えられよう。

また、図 3 から、.60 条件と .70 条件では、被験者が見積った確信度が理論値を上回り、.80 条件と .90 条件では下回っていることが分かる。被験者は低い正答率しか期待できない予測判断課題では過剰確信を、高い正答率が期待できる予測判断課題では過小確信を抱く傾向があると言えよう。この結果は、人が見積る確信度は妥当なものではないとする従来の研究結果を確認するものであるが、本研究の当初の予想とは異なるものである。

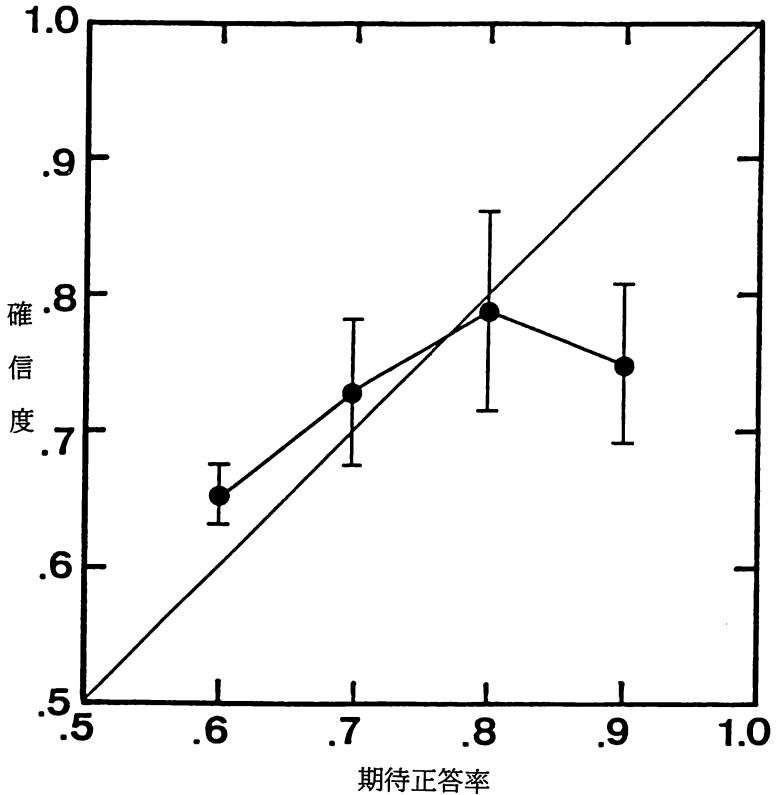


図3 期待正答率の関数としての確信度（傾き45度の直線は理想的な確信度の変化を示す）

4. 全体的考察

被験者の反応確率および被験者が実際に経験する正答率が、理論値によく適合したことから、被験者は与えられた二者択一予測判断課題の事態をかなり正確に把握していた。一方、被験者の確信度は二者択一予測判断課題の事態の変化に応じて単調に推移しなかった。このことから、事態把握過程と確信度形成過程とは異なるメカニズムを持った別の過程考えられる。また、最頻事象の反応確率および生起確率の見積りが確率対応現象を生じるという従来の研究結果から、事態把握過程は試行系列早期に完成すると考えられる。確

信度についても、本実験で試行系列早期に安定することが示唆された。そうであれば、事態把握過程と確信度形成過程とは同時に平行して進行する別の過程であると言える。

確信度形成過程のメカニズムおよび確信度形成過程と事態把握過程との関係が今後検討されなければならない。

要 約

不確定な場面において判断を行うとき、人はその判断に対してさまざまな度合の確信を抱く。人が抱く確信の妥当性は、数値で見積られた確信の度合（確信度）とその確信度が見積られた判断の正答率とを比較することによって評価できる。従来の研究では、人が見積る確信度とその確信度が見積られた判断の正答率とは一致せず、総じて確信度が正答率を上回る過剰確信であって、特に高い確信度が見積られたときほど顕著な過剰確信となる、という結果が得られている。このことから、高い正答率が期待できる判断課題と低い正答率しか期待できない判断課題では、人が自分の判断に対して見積る確信度も異なると考えられる。そこで、本研究では、二者択一予測判断課題を用いて、期待正答率が確信度に及ばず効果を検討した。

得られた結果は次のとおりであった。

1. 被検者の反応確率および実際の正答率は、期待正答率が高くなるに連れて増大した。反応確率は最頻事象の生起確率と極めてよく一致し、実際の正答率は常に期待正答率を下回ったが理論値によく適合した。

2. 第30試行目以降の確信度には、変化・修正が認められなかった。

3. 期待正答率が低い予測判断課題での確信度は相対的に低く、期待正答率が高い予測判断課題での確信度は相対的に高かったが、期待正答率と確信度の間に直線的関係は認められなかった。また、期待正答率が低い課題では確信度は期待正答率を上回り、期待正答率が高い課題では逆に下回った。

以上の結果から、(1) 被験者は期待正答率の違いをかなり正確に識別し、

判断課題を構成する 2 事象の生起確率の比率と同じ比率で予測を行う。(2) 確信度の見積は、一連の試行系列の早期に形成される。(3) 期待正答率が高い課題では確信度の見積が抑制される、ことが指摘された。

さらに、被験者の予測判断課題の事態把握がかなり正確なのに対し、確信度の見積は予測判断課題の事態の変化に単調に対応しないことから、確信度形成過程と事態把握過程は並列に進行する別過程であることが示唆された。

引用文献

- Adams, J. K. 1957 A confidence scale defined in terms of expected percentages. *American Journal of Psychology*, 70, 432-436.
- Adams, J. K., & Adams, P. A. 1961 Realism of confidence judgments. *Psychological Review*, 68, 33-45.
- Adams, P. A., & Adams, J. K. 1958 Training in confidence-judgments. *American Journal of Psychology*, 71, 747-751.
- Beach, L. R., Rose, R. M., Sayeki, Y., Wise, J. A., & Carter, W. B. 1970 Probability learning: Response proportions and verbal estimates. *Journal of Experimental Psychology*, 86, 165-170.
- Beach, L. R., & Swensson, R. G. 1967 Instructions about randomness and run dependency in two-choice learning. *Journal of Experimental Psychology*, 75, 279-282.
- Edwards, W. 1961 Probability learning in 1,000 trials. *Journal of Experimental Psychology*, 62, 385-394.
- Fischhoff, B., Slovic, P., & Lichtenstein, S. 1977 Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 552-564.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. 1980 Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107-118.
- Lichtenstein, S., & Fischhoff, B. 1977 Do those who know more also know more about how much they know? The calibration of probability judgments. *Organizational Behavior and Human Performance*, 3, 552-564.
- Neimark, E. D., & Shuford, E. H. 1959 Comparison of predictions and estimates in a probability learning situation. *Journal of Experimental Psychology*, 57, 294-298.
- Peterson, C. R., & Ulehla, Z. J. 1965 Sequential patterns and maximizing. *Journal of Experimental Psychology*, 69, 1-4.

- Rose, R. M., & Vitz, P. C. 1966 The role of runs in probability learning. *Journal of Experimental Psychology*, 72, 751-760.
- Tversky, A., & Kahneman, D. 1971 Belief in the law of small numbers. *Psychological Bulletin*, 76, 105-110.