

自分自身に関する尺度評定と確信度の関係から見た評定方式の検討

妻 藤 真 彦

美作大学・美作大学短期大学部紀要（通巻第56号抜刷）

論 文

自分自身に関する尺度評定と確信度の関係から見た評定方式の検討

A statistical relation between scale rating of self knowledge and confidence rating.

妻 藤 真 彦

キーワード：自己に関する知識、尺度評定、確信度

質問紙などで広く使われている尺度評定について、これまでその性質・評定過程について詳細な検討が行われてきたが（e.g., Dawson & Brinker, 1971; Dawson & Mirando, 1976; 織田, 1978; Parducci, 1965; Parducci, 1982; Petrov & Anderson, 2005; 椎名, 2008; 脇田, 2004; Wedell & Parducci, 1988; for review, 妻藤, 2006）、そこに“どのようなタイプの情報”が入力されるのか、さらに、入力される情報の種類によってどのように結果が変化するのは、また別のテーマである。

この問題に関連がある研究として、以下のようなものを挙げてよいと思われる：例えば、Klein, Loftus, Traflet, & Fuhrman (1992) は、反応時間を従属変数とした心理実験の結果から、初めて経験する領域での自己概念形成段階では意識的なエピソード記憶が必要とされるが、確定してくると特定のエピソードは不要になると主張した。Lieberman, Jarcho, & Satpute (2004) は、この説を拡張し、自己概念（自己に関する知識：たとえば自分は何が得意で何が苦手かなど）には、“証拠ベース”と“直感ベース”の（意識的にアクセス可能な）2種類があるとし、機能的MRIを用いて2種の自己知識が脳の異なる部位と対応することを確認した。証拠ベースの自己知識はエピソード記憶などの宣言的記憶（declarative memory）に基づくもの

であり、これは顕在記憶（explicit memory）であるから、想起に基づく意識的判断（制御的認知過程）の結果が回答されるため、心理学的には長い判断時間が特徴である。他方、直感ベースの方は判断が自動的であるため、判断時間は前者よりも明確に短く、しかも学習過程はかなりの回数の繰り返しによるので、条件付けのような学習経過をたどると想定される。ただし意識的になされる質問紙への回答に影響しているので、オペラントや古典的条件付けのような潜在（implicit）過程ではないと議論されている。

また行動の実態に関する社会調査について、Mills, Reyna, & Estrada (2006) の研究は、矛盾する調査結果の問題を解決するために、認知・記憶理論を利用できることを示している。それまで思春期のリスクテイキングについて、リスクが大きいと知っているほど実際のリスク行動が少ないという調査結果と、逆にかえって多くなるという調査結果が両方あった。この問題に対して、この研究では質問の仕方によって想起されるエピソード記憶が異なるタイプ（gistあるいはverbatim）になるためではないかという仮説を検討し、質問の仕方によって予測どおりの相反する結果が得られることを確認している。“○○の危険行動を；しばしば行なう、たまに行なう・・・”というような質問・選択肢ではgist想起を誘発し、“自分はそのよ

うな危険行動をいつも行っているわけではない”という、正しいとは限らないエピソード要約内容を答える人が多いのに対し、具体的に“この一月に何回それを行なったか”という質問では、verbatimエピソードの想起によって、gistとは相反する回答になる人が相当の人数になるのである。

これらに加えて、そもそも評定が可能ではない場合があるのではないかという問題提起を、妻藤（2007）は行っている。つまり、ある質問が例えば5点尺度での評定を要求するものであっても、評定過程に入力される情報が不足しているために5段階もの区別ができなかったり、あるいはそもそも2値判断（はい、いいえ）しかできない内容であったときに、どのようなことが起こるかという問題提起である。それでも5点の尺度評定が要求されたとき、2値判断に対する確信の程度を変換して、5段階の評定値を答えるかもしれない。妻藤（2007）はこれが起こっていたかどうかを検出する方法として、各質問項目内で評定値と確信度の個人に渡る相関を計算し、この相関係数が各質問項目の平均評定値に対して1に近い相関関係になるかどうかを検討すればよいと主張した。

例えば評定も確信度も5点尺度の場合、実際に可能なのは2値判断であるから、5点尺度の両極である評定値の5と1は確信が強いときに回答されるはずである。そして、確信がそれよりも弱いときは、4または2が回答され、確信がないときは3が評定値として採用されるはずである（これを確信度変換モデルとする）。このモデルが当てはまるのであれば、評定値から3を減じて絶対値をとると、その値に対して確信度は線型関数になる（1に近い相関であり以下では、これをV字相関とする）。ただし、この関係を検討するとき個人内で項目に渡るV字相関係数をとって平均したり、項目ごとの個人間平均の間のV字相関係数を計算したのでは目的を達成できない。なぜなら、各項目内で確信度から“評定値-3”が決まるとしても、質問項目によって各評定値に対応する確信の強度の範囲が異なっていると（heterogeneous）、これによって生じる項目間分散のせいで、V字相関係数は1から遠い

値になってしまうからである。そこで、次のように別の尺度を工夫する必要がある。

各項目内で評定値と確信度の（V字ではない）個人間相関を計算した場合、評定値の個人間平均が大きい項目であれば、評定値3よりも大きな値が大半であるため、この相関係数は正の値を示すはずである。それに対して平均評定値が3の周辺であれば、3より大きい値と、より小さい値が半々になっているはずであるから、この項目内個人間相関は0に近い値になる。そして、平均評定値が小さい項目では、項目内個人間相関は負の値になる。すると、平均評定値に対して項目内個人間相関をプロットすれば、単調増加関数になるはずである。項目内個人間相関は各項目内で標準化されて直線への当てはまりの程度を表すので、各評定値に対応する確信の強さと範囲が項目ごとに異なっても問題はない。

ただし、このような平均評定値と項目内個人間相関係数の間の相関（以下ではメタV字相関とする）が、ある程度大きくなるもう1つのモデルがある。つまり評定の困難さの相違が存在し、両極の値になるときは評定が容易であるために確信度が高くなるような影響があり、中央に近いほど識別の困難さのために確信度が小さくなるという可能性である。これを評定確信度モデルとする。このモデルの場合、確信度変換モデルと同様に各質問項目の内容による確信度の違いと、それに加えて評定の容易さに基づく確信の違いという2つの要因が確信度に影響する。これらの2つの変動因は互いに独立であるため、メタV字相関は確信度変換モデルよりも小さくなる（妻藤，2009）。

妻藤（2009）のシミュレーションによると、確信度変換モデルが成立しているときのメタV字相関は0.9以上であり、ほぼ0.95以上であったが、評定値と確信度が独立の場合には-0.4から+0.45の範囲になった（妻藤，2009）。そこで、実データが0.5から0.9の範囲に入っていたときは、評定確信度モデルが当てはまり、0.95以上であれば確信度変換モデルが当てはまる可能性が高いと解釈される。各々の間にグレーゾーンがあるという問題は残っているが、以下のように、こ

れらの大雑把な基準でも、妻藤（2007）の結果に含まれる3つの条件の相違を識別できると解釈された。確信度自体がどのように決定されるかという理論的問題はまだ十分な決着にいたっていないが（妻藤, 1992; Saito, 1998; 妻藤, 2004; Koriat, 2008; 妻藤, 2010）、ここでの文脈では確信度自体に関する評定過程は考慮しなくても、以上の考察に問題は生じない（妻藤, 2009）。

妻藤（2007）では、他者の行動がどの程度意図的であったかという判断の性質を検討するために Malle & Knobe (1997) が作成した質問セットを用い、メタV字相関を計算した。何の理由も原因も記述されず単に行動のみが呈示される条件（例えば、“彼女はスーパーに買い物に行く途中、スピード違反をした”）で、メタV字相関は0.97、行動の内的理由が付加されていた条件（例：“彼女はスーパーに買い物に行く途中、急いでいたのでスピード違反をした”）では0.79、行動の（外的）原因が付加された条件（例：“彼女はスーパーに買い物に行く途中、閉店時間が迫っていたのでスピード違反をした”）では、0.91であった。そうなると、単に行動のみを述べた条件では確信度変換モデル、理由条件では評定確信度モデルが当てはまると言えよう。ただ原因条件の場合は、グレーゾーンであって明確でないということになる。

このように行動のみと理由付の相違は、他者行動がどの程度意図的であったかを推測するときに、必要な情報が不足していたり、2値判断しかできない内容であったときに生じるという仮説に当てはまるものである。内的理由が書かれている場合、意図的であったかどうかの推測と直結するが、外的原因の場合は、それに基づいて内的理由を推測し、さらに意図性を推測することになるため、回答者によって評定確信度モデルの方式になったり確信度変換モデル方式になったりして、メタV字相関がグレーゾーンになるのであろう。

このように、5段階評定を求められたとき、意図的かそうでないかという2値判断しかできない場合には、何とか評定値を決めようとして、おそらく非意識的に（自動処理的に）確信度を利用して程度の相違を

作り出していると思われる。

この実験では他者の意図性を推測するという課題であり、与えられた情報の量と性質によって評定値を決定する方法が変化すると解釈された。では他者に関する推測ではなく自分自身に関する評定の場合、情報の量あるいは情報の信頼性が異なっているときにはどうなるかも検討しておくべきであろう。本稿の目的は自分自身に関する情報の量や信頼性の違いを設定したとき、メタV字相関がどのくらいの値になるかを調べることである。

具体的には、現在の自分に関する評定と、過去の自分に関して記憶に頼って同じ評定を行い、それらを比較するという方法を用いる。過去の自分に関する記憶は、現在の自分に関する情報よりも少なかったり、記憶内容に関する確信の減少等が生じているはずである。ただし、その現在との相違が小さすぎると意図性の推測ほどメタV字相関の違いが出ないかもしれないので、大学生を回答者として、高校生の時の記憶と、さらに過去である中学生の時の記憶の両方を用いて、その間の相違も検討する。

方 法

回答者 大学1年生および高等専門学校4年生（大学1年生と同じ年齢）161名（女性89、男性72）。ただし回答に不備があった6名が除外され、さらに“中学”条件が最も少なく51名となったため、“現在”条件で2名、“高校”条件で1名をランダムに除き、各条件51名となるようにされたので、最終的に153名のデータが分析された。各条件で人数を揃えたのは、質問項目の相違を見るための分散分析について、計算精度が低くならないようにするためである。

質問紙 近藤・鎌田（1998）が作成した“生きがい感スケール”の中から“現状満足度”“人生享楽”“意欲”の3つの下位尺度項目20個が使用された。このスケールにはもうひとつ“存在価値”も含まれているが、この下位尺度は“私は他人から信頼され頼りにされています”というような一種の“価値評価”およ

び“他者が感じている評価の推測”を伴う質問から構成されており、他の下位尺度の質問と異なる側面を持っているため、以下の質問紙では使用されなかった。このようにして、使用された質問項目はすべて自分の行動や感じている内容に関するものになるようにされた（例えば“今日は一日好きなことができると思う日がよくあります”など）。このようにしたのは、性質の異なるものが混ざることによって、評定値と確信度の関係が複雑になる可能性を避けるためである。

この20個の“生きがい感”評定は、もとのスケールでは3点尺度であるが、ここでは5点尺度とし、スケール上に○を記入することで行われた。またそれら各々の評定に対する確信度も5点尺度とし、点数を数値でカッコ内に記入する形式とされた。このように両評定の回答形式を変えたのは、妻藤（2007）と同様、“生きがい感”と確信度の回答パターンが視覚的に比較できないようにするためである。

この質問紙（“現在”条件用）の他に、各文に“高2のころ”あるいは“中2のころ”が付加され、文末を過去形に変えた質問紙（各々“高校”条件用、“中

学”条件用）も作られた。この3条件の違いを“記憶条件”とする。これらの各々について、質問項目を異なるランダム順に並べた質問紙が2種作成され、6種類の質問紙がランダム配布された。したがって記憶3条件は異なる回答者群間比較である。また1つの条件について質問のランダム順が2通り作成されたのは、各条件内での質問項目呈示順をカウンターバランスするためである。さらに3条件が異なる被験者群とされたのは、もし3条件の全てを一人の回答者に呈示した場合、確信度の評定も含むために、回答数が120となって回答時間が相当長くなるのを避けるためである。また同じ内容の質問に3回も答えることになる、要求特性によって結果が歪む可能性もあると思われるためである。

結果と考察

図1に示すように質問項目間にはかなり大きな平均値の相違があるが、記憶条件による違いは項目によって異なっている。実際、3記憶条件（個人間）×20

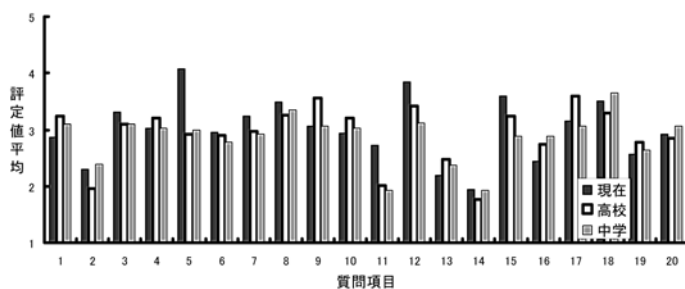


図1 生きがい感平均

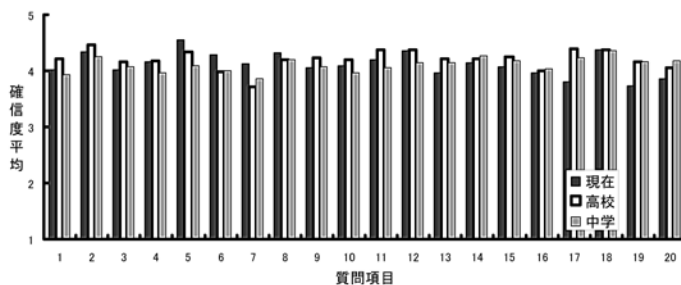


図2 確信度平均

質問項目（個人内）の分散分析によると、記憶条件の主効果は有意ではなく、質問項目の主効果は有意である ($F(19, 2850)=34.37, p < .001$)。また記憶条件×質問項目の交互作用も有意であった ($F(38, 2850)=3.61, p < .001$)。ただし球面性が有意に破れていたため (Mauchly検定)、自由度調整法によって補正も行なったが、結果は変わらなかった。このように、後の相関分析を行う前提として必要な、質問項目間の有意な相違が確認された。

図2に確信度平均を示す。これについても質問項目間に平均の差が見られ、記憶条件による違いは項目によって異なっている。分散分析も同様の結果であり、記憶条件の主効果は有意ではなく、質問項目の主効果 ($F(19, 2850)=3.33, p < .001$) と、記憶条件×質問項目の交互作用 ($F(38, 2850)=1.51, p < .05$) が有意であった。これも球面性の破れを考慮した自由度補正を行っても結果は変わらなかった。

このように以下の分析に必要な前提が満たされたので、メタV字相関係数が計算された (表1に示す)。どの条件も確信度変換モデルのカットオフである0.95を越えていない。また独立モデルのカットオフである0.45よりも大きいので、3条件とも評定確信度モデルが当てはまることになる。

表1 記憶条件ごとのメタV字相関係数

| 現在 | 高校 | 中学 |
|------|------|------|
| 0.79 | 0.87 | 0.86 |

仮説としては評定過程に入力される情報について、その量と内容の性質によって、評定方式（確信度変換、評定確信度、独立モデル）が変化するというものであったが、今回の結果から見て、情報の量というよりは内容の性質の方が重要なかもしれない。

ただし、今回の条件設定では、方式の違いを誘発するほどの情報量の相違が作れなかったという可能性は、かなり大きく、確信度平均には記憶条件差の主効果がなかったことが、この可能性を支持する結果となっている。図2を詳細に見ると、質問項目によ

て、現在が最も大きくなっているものと、中学が最も大きくなっているものもある。おそらく思春期であり、かつ中学・高校という多くの人が類似した経験をしやすい環境にある年齢の場合、ある時点で多くの人に強い印象や情動を喚起するようなエピソード、あるいは類似した動機付け等によって記憶に残りやすい項目内容があるために、単に過去の記憶であるというだけでは、質問セット全体としての相違を設定できないのであろう。

とはいえ“高校”と“中学”のメタV字相関は0.87と0.86であり、極めて近い値であるのに対して、“現在”は0.79という僅かとも言えない相違がある。このような評定確信度モデル内でのメタV字相関の違いは、現在の理論では扱えない側面を示唆する可能性もあり、自分自身の現在と過去の評定については、まだ相当の検討が必要であろうと思われる。

他方、他者の意図性の結果 (妻藤, 2007) を見ると、ある行動の内的理由が与えられると評定確信度モデルが当てはまるが、行動の (外的) 原因の形である場合は、モデルの識別グレーゾーンになった。この相違は情報量の違いというよりも、内容の性質によると解釈できよう。序論で論じたように、内的理由は意図性の推測と直結するのに対し、外的原因は、そこからまず内的理由を推測し、その後在意図性の推定に進む可能性がある。そうだとするならば、これは情報量の問題ではなく、内容の性質に関わる相違である。

今回のような、自分自身に関する評定の記憶条件と質問項目の交互作用が、同質の情報に関して記憶の曖昧さ等の違いによると考えられるなら、評定方式は入力情報の内容の性質によって切り替わると考える方が理論的に整合的である。これを確認するために、今回のような条件設定ではなく、より明確に情報量を操作できる設定を案出することが今後の課題である。

結論と要約

自分自身や他者に関する尺度評定について、2値判断しかできない場合に、多段階尺度が強制されると確

信度が変換されて回答されるというモデルと、評定値の両極は判断が容易であることによって、結果的に確信度との統計的關係が生じるという2つの理論的モデルが、自分自身の内的な側面に関する評定を現在と過去に関して行なった結果から検討された。その結果、このような回答の場合は、確信度変換モデルではなく、評定値から確信度への影響が出るというモデルの方が良い当てはまりを示した。ただし、そのような原因については、今回のような現在と過去の相違だけでは、記憶の記録時の条件を統制できないという問題があり、情報の量の違いが原因なのか内容の性質の違いなのか、明確な結論には至らなかった。とはいえ、現在の自分に関する評定で確信度変換モデルが当てはまらないということは、そのような質問紙の場合、2値判断ではなく、多段階評定尺度を用いる方がより妥当であることを示している。

引用文献

- Dawson, W.E., & Brinker, R.P. (1971). Validation of ratio scales of opinion by multimodality matching. *Perception & Psychophysics*, 9, 413-417.
- Dawson, W.E., & Mirando, M.A. (1976). Inverse scales of opinion obtained by sensory-modality matching. *Perceptual and Motor Skills*, 42, 413-425
- 近藤勉・鎌田次郎 (1998). 現代大学生の生きがい感とスケール作成. *健康心理学研究*, 11, 73-82.
- Klein, S. B., Loftus, J., Trafton, J. G., & Fuhrman, R. W. (1992). Use of exemplars and abstractions in trait judgments: A model of trait knowledge about the self and others. *Journal of Personality and Social Psychology*, 63, 739-753
- Koriat, A (2008). Subjective confidence in one's answers: The consensus principle. *Journal of Experimental Psychology: Learning Memory and Cognition*, 34, 945-959.
- Liberman, M.D., Jarcho, J.M., & Satpute, A.B. (2004). Evidence-based and intuition-based self-knowledge: An fMRI study. *Journal of Personality and Social Psychology*, 87, 421-435.
- Malle, B.F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33, 101-121.
- Mills, B, Reyna, V.E., & Estrada, S. (2006). Explaining contradictory relations between risk perception and risk taking. *Psychological Science*, 19, 429-433.
- 織田輝準 (1978) 評定尺度による判断過程の研究 *教育心理学研究*, 26, 142-151
- Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, 72, 407-418.
- Parducci, A. (1982). Category ratings: Still more contextual effects! In B. Wegener (Ed), *Social attitudes and psychophysical measurement*. Hillsdale: Lawrence Erlbaum. Pp.89-105.
- Petrov, A.A. & Anderson, J.R. (2005). The dynamics of scaling: A memory-based anchor model of category rating and absolute identification *Psychological Review*, 112, 383-416.
- 妻藤真彦 (1992). 根拠を述べるができない確信と「意識様態」. 美作女子大学・美作女子大学短期大学部紀要, 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 (2004). 確信度評定のメカニズムと理論的問題. 風間書房
- 妻藤真彦 (2006). 尺度評定過程への入力情報の問題. 美作大学・美作大学短期大学部紀要, 51, 1-10
- 妻藤真彦 (2007). 質問紙評定過程における参照情報－他者行動の評定－. *心理学研究*, 77, 541-546
- 妻藤真彦 (2009). 尺度評定と評定の確信度との関係：質問セットによる相違のシミュレーション. 美作大学・美作大学短期大学部紀要, 54, 19-28.
- 妻藤真彦 (2010). 確信度評定過程におけるサンプリング理論の比較：繰り返し質問における解答と確信度の変化. 美作大学・美作大学短期大学部紀要, 55, 49-57.
- 椎名乾平 (2008). カーソルの軌跡を用いて評定判断過程を分析する. 早稲田大学教育学部：学術研究——教育心理学編——, 56, 1-9.
- 脇田貴文 (2004). 評定尺度法におけるカテゴリ間の間隔について. *心理学研究*, 75, 331-338.
- Wedell, D.H., & Parducci, A. (1988). The category effect in social judgment: Experimental ratings of happiness. *Journal of Personality and Social Psychology*, 55, 341-356.