

Title	確立を用いた推論課題における回答方略の検討
Author(s)	堀下, 智子
Citation	大阪大学大学院人間科学研究科紀要. 31 P.11-P.24
Issue Date	2005-02
Text Version	publisher
URL	https://doi.org/10.18910/3591
DOI	10.18910/3591
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/repo/ouka/all/>

確率を用いた推論課題における回答方略の検討

堀 下 智 子

目 次

1. はじめに
2. 方 法
3. 結 果
4. 考 察

確率を用いた推論課題における回答方略の検討

堀下 智子

1. はじめに

推論とは、情報や知識をもとにして新しい結論を導き出す過程を指す。例えば、外出時に傘を持っていくかどうかを決定する、あるいは、目的地に早く到着するためのルートを考える、などの過程がそうである。

推論に用いられる情報は、多くの場合、確実ではない。例えば、天気予報から得られる降水確率や、高速道路が混雑する見通しなど、我々を取り巻く情報に「100%確実である」ものはむしろ稀である。このような情報の不確実性を確率として表現し、その確率情報から何らかの結論を導き出す過程のことを特に、確率推論とよぶ。確率推論は、前述のような意思決定場面、さらに出来事の評価・解釈に関わるという点で、人間の基礎的な認知的機能の一つであると考えられている(南, 1994)。

我々の日常のありふれた行動は、このような推論の積み重ねであるといっても過言ではない。しかし、Kahneman, Slovic & Tversky (1982)などのこれまでの研究から、確率を用いた推論課題において人間がしばしば規範解から逸脱した行動をとるという現象が報告されている。

本研究は、確率を用いた推論課題における人間の規範解からの逸脱について論じ、さらに、推論結果だけでなく、課題解決の過程で行われる実際の行動について明らかにすることを目的とする。

基礎比率の無視 Tversky & Kahneman (1973)は、確率推論における規範解からの逸脱の例を多く挙げ、指摘している。その中でも特に有名であり、現在でも研究が行われているのが、「基礎比率の無視」という現象である。

以下の「感染者問題」と呼ばれる課題を用いて、この現象を説明する。

【代表的な感染者問題】(Eddy, 1982)

乳がんの発症率は、1%である。

患者が乳がんを患っている場合、検査で陽性反応が出る確率は79%である。

患者が乳がんを患っていない場合、陽性反応が出る確率は9.6%である。

ある患者に陽性反応が出た場合、実際に乳がんを患っている確率はどの程度か？

この課題では、基礎比率は「乳がんを患っている」=1%である。この確率と、79%と9.6%という二つの条件つき確率によって、「陽性反応が出た」という状況下での「乳がんを患っている確率」を回答する事が求められる。ベイズの定理に基づく正答は以下の通りである。

$$\frac{0.01 \times 0.79}{0.01 \times 0.79 + 0.99 \times 0.096} = 0.077 \quad (7.7\%)$$

しかし、多くの実験協力者の直感的な判断は、75%前後である。この課題では、実際に感染している人の割合が1%と非常に低い。また、感染していない人の中にも陽性反応の出る人が存在する可能性を考慮する必要がある。しかし、実験協力者の多くはこの基礎比率(1%)の情報を軽視し、回答が条件つき確率(乳がんを患っている場合に陽性反応が出る確率:79%)に引きずられ、75%前後という回答を導き出すものと考えられる。

Tversky & Kahneman (1973)は、ヒューリスティックスという推論ルールを提唱した。ヒューリスティックスとは、人が問題解決、判断、意思決定をおこなう際にしばしば用いられる、必ずしも解決の保証は無いが、うまくいけば解決できる可能性が高い、比較的簡単な方略のことである。「基礎比率の無視」などの判断の逸脱は、ヒューリスティックスの適用失敗によって生じた現象であると説明している。

しかしながら、ヒューリスティックス説は現象を単に「ヒューリスティックス」という言葉によって定義づけたに過ぎず、プロセスに言及していない。また、なぜ、どのように間違いを犯すのか、どのようにすればその間違いが防げるのか、という事についても解決することが出来ていないという点で、多くの批判がある。

確率表現形式に関する議論 前述のような確率推論の誤りに関して、最近、さまざまな研究が報告されており、特に重要視されているのが確率の表現形式である。

確率推論の課題で確率を表すためには、「20%」というように%を用いたもの(以下、確率形式)、標本抽出の形で「10回のうちの2回」というように表したもの(以下、頻度形式)、そして「時々」や「おそらく」というように頻度語を用いたもの(言語確率)、などの方法が用いられる。言語確率は確率・頻度形式のような数量的表現に比べて意味に幅があり、文脈や個人間の差も大きいと考えられる。確率形式と頻度形式では、異なる表現で同じ確率を表す事が可能であるが、その違いこそが、確率推論課題の困難さを左右するのではないかと指摘されている。

Cosmides & Tooby (1996)の実験では、前述の感染者問題を以下のように改変した課題を用いた。この課題は、感染者問題における確率情報をすべて頻度に置き換えたものである。

【感染者問題の頻度版】

1000人の女性のうち、乳がんを発症するのは10名である。

10名の患者のうち、検査で陽性反応が出るのは8名である。

990名の発症していない人のうち、陽性反応が出る確率は96名である。

陽性反応が出た女性のうち、実際に乳がんを発症しているのは何人中の何人か？

このような課題を用い、確率情報が確率形式で与えられた場合と頻度形式で与えられた場合とで比較を行った。その結果、頻度形式で与えられた場合には、確率形式で与えられた場合に比べて、より良い成績、すなわちベイズ定理に基づいた規範解を出す事が示された。同様に、Gigerenzer & Hoffrage (1995)も、問題が頻度形式で呈示された場合は、確率形式で呈示された場合よりも、推論の誤りが少なくなったと報告した。

この結果から、Cosmides & Tooby (1996)は、人間の確率推論のメカニズムが、「加工された頻度情報」、すなわち頻度形式に合致するように進化したのではないかと主張している。また、Gigerenzer & Hoffrage (1995)らは、経験を基礎とした頻度としての確率表現の方が、人間は理解しやすいのではないかと考察した。ここで彼らの主張の中核をなすのは「モジュール説」(Gigerenzer, 1995; Gigerenzer & Hoffrage, 1995など)である。彼らによると、採集狩猟生活者として生きていた先祖は、環境中の出来事の頻度を観察することはできる。そして、過去に経験した出来事と同じ状況で再度意思決定を行う場面において、その場面に関する情報を再生することができることは、非常に適応性のあることである。このことから、彼らは、「頻度情報が入力された」という文脈においてのみ働く確率推論のモジュールという、適応性のあるメカニズムを人間が持っている」と結論づけている。

このように、Gigerenzer & Hoffrage (1995)らの主張は一見、これまでのヒューリスティックス議論に明確な結論を与えたかに見える。すなわち、確率課題で実験協力者が正答を導き出せないのは、%という確率表現形式が人間にとって適応的でないからである、という主張である。しかしながらこの主張は、%を用いた課題解決場面において、基礎比率を無視するという現象がなぜ起こるのか、という根本的な原因を説明するものではなく、単に頻度表現の優位性を説明する一つの主張に過ぎない。

モジュール説への批判は、彼らの行った他の研究に対しても向けられている。例えば、Gigerenzer (1998)は、確率情報を頻度表現に変換し、ベイズ推論を正しく行えるようにすることを目的とした訓練を開発することを試みた。その結果、確率表現のままの情報をベイズ推論に用いる方法を訓練した群よりも、頻度情報をベイズ推論に用いる事を訓練した群の方が、課題の正答率が上昇するという結果が得られた。しかし、このような訓練によって頻度情報の利用を学習することができるという主張は、頻度形式が人間にとって適応的であるという彼らの主張と矛盾する。

また、Gigerenzer & Hoffrage (1995)の実験では、訓練の効果を5週間後に確認し

たに過ぎず、このような訓練の効果があつたと主張する他の研究(Fong & Nisbett, 1991)も、同程度の期間の効果しか追跡できていない。さらに、Reeves & Weisberg (1993)は、長期間経った後の追跡調査では、このような訓練の効果が認められないことを示している。頻度情報の効果が適応性を示すものならば、訓練の効果は永続的であるべきであり、これらの研究結果も、モジュール説の主張と矛盾する。

本実験の目的 これまで述べたように、感染者問題に代表される確率推論の過程には、さまざまな要因が横たわり、課題解決を困難にしていることがうかがえる。そこで、本実験は、そのような確率推論の困難さの原因を明らかにする事によって、確率推論課題における推論過程を検討した。

まず、確率を用いた推論課題の困難さの原因を明らかにするため、課題解決のプロセスを細分化することによってパフォーマンスを検討する。特に、本実験では、条件つき確率（ある条件下においてある事象が生起する確率）の理解に焦点を当てる。従来、頻度表現を用いた課題では回答を「何人中の何人」で求めるのに対し、%を用いた課題では条件つき確率を%で回答させる。つまり、頻度課題では条件つき確率の意味が回答の形式として分かりやすく呈示されているのに対して、%課題ではそのような助けなしに条件つき確率を求めさせているのである。この方法では、確率表現の違いだけでなく、課題構造の違いがパフォーマンスに影響を与えている可能性がある。そこで、本実験では課題の回答方式を二段階とし、まず「何人中の何人」の形式での回答を求め、次に%形式で条件つき確率を回答させるという形式に統一した。もし、従来の%形式課題におけるパフォーマンスの低さが条件つき確率の理解のし難さによるものならば、「何人中の何人」の質問を挿入することによって、頻度課題と変わらない正答率が得られると予想される。なお、本実験では条件つき確率の理解に焦点を置くため、従来指摘されている基礎比率については、確率・頻度どちらの課題でも頻度で提示するように統一した。

また、確率表現の違いを原因とするモジュール説では頻度表現の優位性を主張するにとどまっており、この説には批判も多い。もし頻度表現の優位性がモジュールによるものならば、この優位性は課題内容に関わらず確認されるはずである。そこで本実験では、同じ確率情報を用いているが内容の異なる二種類の課題を用意し、そのパフォーマンスを比較する。

2. 方 法

実験計画 確率形式(確率/頻度)×課題内容(感染者/ニュートラル)×質問(質問1/質問2)の3要因混合計画であった。確率形式要因、課題内容要因は被験者間要因、質問要因は被験者内要因であった。

実験協力者 大学生143名(平均年齢20.2才)であった(男性:79名, 女性:64名)。感

染者／確率条件に 35 名、感染者／頻度条件に 35 名、ニュートラル／確率条件に 37 名、ニュートラル／頻度条件に 36 名がそれぞれ割り当てられた。

材料 課題は、Kahneman, Slovic & Tversky (1982)を参考に作成された。感染者条件、ニュートラル条件の課題文を、以下に示す。また、課題文中の確率情報は、確率条件、頻度条件ともに、基礎比率(1000 人中の 40 人)は頻度で提示され、その後の情報は確率条件では確率(%)で、頻度条件では頻度で提示された。また、質問は 2 問用意された。質問の内容は確率形式の条件に関わらず同じものであった。

【感染者／確率条件】

問題

ある感染症に対する検診方法が研究されてきました。
ここに、その感染症に関する情報と検査結果が示されています。
1000 人のうち 40 人が、この感染症に感染しています。
もし感染していた場合、この診断方法を利用すると 75%の確率で「陽性反応」が出ます。
しかし、もし感染していない場合にも、12.5%の確率で「陽性反応」が出てしまいます。
(頻度条件では、問題の最後二文が以下のように異なった。)
40 人の感染している人のうち、30 人は、「陽性である」と診断されます。
960 人の感染していない人のうち、120 人は「陽性である」と診断されてしまいます。

質問 1

この中で、診断の結果、陽性反応が出る人のうち、実際に感染しているのは何人でしょうか？

質問 2

いま、Aさんが検診を受けました。もしAさんに「陽性反応」が出た場合、実際にAさんがこの感染症に感染している確率は何%でしょうか？

【ニュートラル／確率条件】

問題

広場に人が 1000 人集まっています。
このうちの 40 人が、メガネをかけています。
メガネをかけている人のうち、75%の人が、帽子をかぶっています。
メガネをかけていない人のうち、12.5%の人が、帽子をかぶっています。
(頻度条件では、問題の最後二文が以下のように異なった。)
メガネをかけている 40 人のうち、30 人が帽子をかぶっています。

メガネをかけていない960人のうち、120人が帽子をかぶっています。

質問1

この中で、帽子をかぶっている人のうち、メガネをかけているのは何人でしょうか？

質問2

いま、この中の一人「Aさん」が帽子をかぶっていることがわかっています。

さて、この「Aさん」がメガネをかけている確率は何%でしょうか？

手続き 実験は、近畿圏内の私立大学において「認知科学」および「行動科学」の授業時間内に集団実験の形で行われた。課題は冊子に印刷された。課題の一ページ目には課題文と解答欄が印刷された。条件間で課題文が異なったため、課題文を実験者が読み上げず、課題解決に用いた図表や計算式など筋道を示すものは全て質問紙中に残すように指示された上で、6分間の回答時間内に何らかの答えを出すように求めた。実験協力者の回答は、「質問1」では「□人のうちの□人」、「質問2」では「□%」というそれぞれの□の中に当てはまると思う数字を記入する方法で行われた。課題の二ページ目では、自分が回答した答えに対する確信度と、課題回答時のイメージ化の程度、そして課題文の理解しやすさに関して、5段階で回答を求めた。

実験の最後には、本実験で用いられた課題およびそれに類似した課題を授業等で経験したことがあるかどうかを調査した。

3. 結 果

正答率 1名の実験協力者が、本実験で用いられた課題を以前に見たことがあると回答したため、分析から除外した。よって、以下の分析はすべて142名の実験協力者をもとに行われた。

どちらの課題内容においても、質問1では「150人のうちの30人」、質問2では「20%」としたものを正答とみなした。

まず、条件ごとの正答率を比較した。条件ごとの正答率を Figure 1 に示す。各条件の正答率をそれぞれ逆正弦変換し、カイ二乗検定を行った。まず、質問1について、2(確率形式/確率・頻度) × 2(課題内容/感染者・ニュートラル) のカイ二乗検定を行ったところ、確率形式 ($\chi^2(1, N = 142) = 5.41, p < .05$) および課題内容 ($\chi^2(1, N = 142) = 5.98, p < .05$) の主効果が有意であり、交互作用は有意ではなかった。また、質問2についても同様の比較を行った結果、課題内容の主効果は有意 ($\chi^2(1, N = 142) = 4.12, p < .05$) であったが、確率形式の主効果および交互作用は有意ではなかった。以上の結果から、どちらの質問においても、ニュートラル条件の正答率が感染者条件の正答率よりも高いことが明らかになり、また質問1においては、頻度条件の正答率が確率条件の正答率よりも高いことが明らかになった。

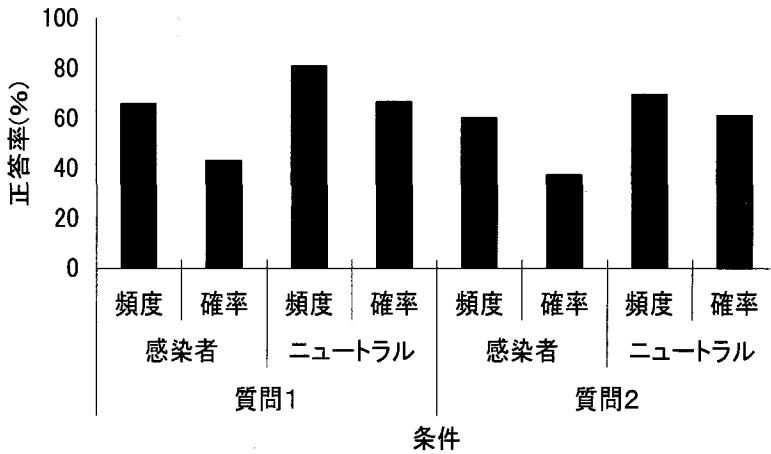


Figure 1 条件ごとの正答率(%)

次に、課題内容ごとに、同様に2（確率形式／確率・頻度）×2（質問／質問1・質問2）のカイ二乗検定を行った。その結果、ニュートラル条件においては確率形式・質問のどちらの主効果も認められず、交互作用も有意ではなかった。一方、感染者条件においては、確率形式($\chi^2(1, N = 70) = 7.58, p < .01$)の主効果のみ有意であった。この結果から、どちらの課題内容においても質問1と質問2の正答率に差は認められず、また感染者条件においてのみ確率条件より頻度条件の方が正答率が高いことが明らかになった。

イメージ化 課題に回答する際にイメージ化を行ったかどうかを5段階評定(5：とても具体的にイメージした～1：まったくイメージしなかった)で回答を求めた。条件ごとの平均値を Figure 2 に示す。

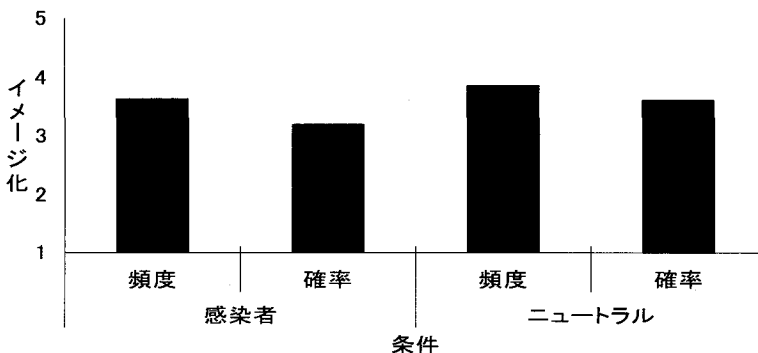


Figure 2 条件ごとのイメージ化の程度

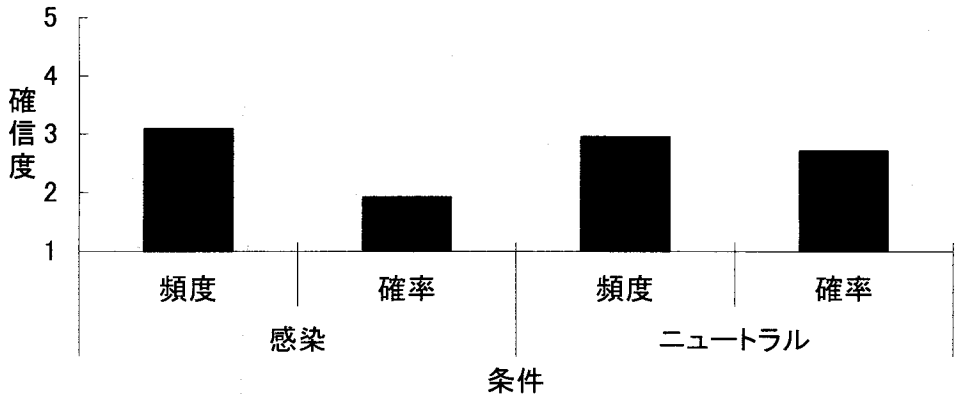


Figure 3 条件ごとの平均確信度

2 (課題内容/感染者・ニュートラル)×2 (確率形式/確率・頻度)の2要因分散分析を行った結果、課題内容の主効果が有意であり($F(1,142) = 4.43, p < .05$), 確率形式の主効果も有意であったが($F(1,142) = 4.92, p < .05$), 交互作用は有意ではなかった。このことから、確率条件よりも頻度条件のほうが評定値が高く、また感染者条件よりもニュートラル条件のほうが評定値が高いことが明らかになった。

確信度 5段階評定によって回答された確信度について分析を行った。条件ごとの平均確信度を、Figure 3に示す。確信度は5段階評定であり、値が大きいほど確信度が高いことを示す。

2 (確率形式/確率・頻度)×2 (課題内容/感染者・ニュートラル)の2要因分散分析を行った結果、確率形式の主効果が有意であった($F(1,142) = 9.00, p < .01$)。課題内容の主効果および交互作用は有意ではなかった。この結果より、確率条件よりも頻度条件のほうが、実験協力者の確信度が高いことが示された。

理解度 課題文および質問の内容についての解りやすさについて、5段階評定によって回答を求めた。この評定値について、2 (確率形式/確率・頻度)×2 (課題内容/感染者・ニュートラル)の2要因分散分析を行った結果、確率形式・課題内容ともに主効果は有意ではなく、交互作用も有意ではなかった。

4. 考 察

本実験は、感染者およびニュートラルという二種類の確率推論課題を用い、質問を二段階に設定し、課題中の確率情報が確率(%)で提示された場合と頻度で提示された場合とでの正答率の比較を行った。まず、質問1と質問2の正答率は、どの条件間でも有意な差が認められなかった。そして、ニュートラル条件は感染者条件より正答率が高いという結果が一貫して得られた。さらに、確率の提示形式は、ニュートラル条件においては正答率に影響を与えなかったものの、感染者条件では確率条件より頻度条件の正答率が高いという結果が得られた。そして、これらの正答率の結果と、確信度およびイメージ化の程度の結果とは、ほぼ連動するものであった。

質問2の正答率は、従来の研究で得られた結果より、高い傾向にあった。質問1と質問2の間に正答率の差が見られなかったことから、質問1で条件つき確率の意味を強調することによって、従来のように質問2の回答を単独で求める課題に比べて質問2の正答を促進したためであると考えられる。また、従来の確率課題は、確率情報が与えられた後で直接、条件つき確率を問う形であったが、頻度課題では条件つき確率を問うのではなく、その意味ともいえる「何人中の何人」を答えさせている。このように実験協力者に要求する回答がそもそも異なっていたのである。本実験で、この質問の形式を統一した結果、最終的な条件つき確率の正答率は頻度条件と確率条件で差が見られなかった。このことから、これまでの研究で指摘されてきた確率推論課題の困難さは、質問形式に依存する可能性が示唆された。これまでの研究では、人間は「確率推論が出来ない」と結論づけられがちであったが、本研究からも明らかになったように、人間は確率を計算・操作する能力を十分に備えており、本実験で示されたような条件が整えば、それを実際の確率推論において遂行することが可能なのである。

また、本実験で導入した課題内容の操作は、これまでの Gigerenzer (1995)などの主張と矛盾するものであることが示された。彼らの主張では、頻度形式の優位性は頻度形式が適応的であるからであるという説明であった。しかし、本実験では、頻度形式の優位性は、感染者条件では認められたが、感染者条件では認められた。すなわち、この優位性は課題内容に依存することが明らかになった。もし、Gigerenzer (1995)などの主張するように、頻度情報が入力された場合に自動的に起動するモジュールが存在するのであれば、課題内容が異なったとしても、回答に差が出ないはずである。よって、以上の結果は、頻度形式の課題は人間にとって解決が容易であるという現象をモジュールの適用という視点から説明することの限界を示している。

さらに、ニュートラル条件は感染者条件より正答率が高いという結果が一貫して得られた。確率形式の優位性だけでなく、課題内容が課題解決を左右する可能性が示唆される。本実験では、いわゆるオリジナルな感染者問題と同じ構造をしていながらも正答率

の異なるニュートラル課題を作成し、確率推論における課題内容の効果を示した。このことは、確率推論研究における方法論の面から、意義のあることであると言えよう。

ただし、課題内容によって課題文の理解度に差が認められなかったことから、少なくとも本実験での内容操作は、課題自体の理解度を左右したとは認められない。すなわち、正答率の低さの原因が課題文の解りにくさにあるわけではなかった。では、この課題内容による回答の差は何によるものなのであろうか。その手がかりの一つが、イメージ化の評定値である。課題解決時におけるイメージ化の程度は、正答率とほぼ連動する結果となっている。すなわち、課題のイメージのし易さが、課題解決を促進している可能性がある。しかし、本実験では、そのイメージ化の内容まで踏み込んだ分析を行うことが出来なかったため、今後はこの点について詳しく確認できる実験操作を行うべきであると考えられる。

イメージ化に関する議論はメンタルモデル理論(Johnson-Laird & Byrne, 1991)にも大きく関連する。メンタルモデルとは、そのときに考えている事柄の内容に応じて心の中に作り出され、利用される表象(高橋・服部, 1996)であり、この理論によると、人間は推論を行う際に、抽象的な論理操作によって推論するのではなく具体的なモデルを心的に操作して推論すると仮定されている。本研究におけるイメージ化とは、実際に何人かの感染者や陽性反応の出た人をイメージし、操作を行うことを指す。そして、本研究のような確率推論において人間は、「感染しているのに陽性反応が出ない」あるいは「感染していないのに陽性反応が出る」といった表象をメンタルモデルとして作り、操作していると推察される。

今後は、イメージ化をはじめとする課題理解の問題を、このような既存の理論との整合性の面からも議論・検討する必要があるだろう。

本実験で得られた結果は、確率推論課題の困難さを説明するモジュール説の限界を示唆するものである。課題の回答は内容・確率表現形式によって左右されることから、確率推論過程に横たわる頑健かつ汎用的なモジュールが存在するとは言いがたいことが明らかになった。また、確率を用いた推論課題の回答が、課題における質問の内容や意味の理解に依存していることが明らかになった。このことから、今後は、課題の回答のみを対照とするのではなく、課題解決のプロセスや理解、さらに解決に関わる表象の内容についても詳細な検討を行うことが求められるであろう。

引用文献

- Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58, 1-73.
- Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities.

- In D. Kahneman, P. Slovic, & A. Tversky (Eds), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General*, 120, 34-45.
- Gigerenzer, G. (1995). The bounded rationality of probabilistic mental models. In K. I. Manktelow & D. E. Over (Eds.), *Rationality*. London: Routledge and Kegan Paul.
- Gigerenzer, G. (1998). Ecological intelligence: An adaptation for frequencies. In D. D. Cummins. & C. Allen. (Eds.), *The evolution of mind*. London: Oxford University Press.
- Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684-704.
- Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S. & Caverni, J. P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review*, 106, 62-88.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- 南学 (1994). 問題解決事態としての確率判断. 広島大学教育学部紀要, 41, 153-160.
- Reeves, L. M., & Weisberg, R. W. (1993). Abstract versus concrete information as the basis for transfer in problem solving: Comment on Fong and Nisbett (1991). *Journal of Experimental Psychology*, 122, 125-128.
- 高橋 和弘・服部 雅史 (1996). 演繹的推論. 市川伸一 (編) 認知心理学4 思考. 東京大学出版会
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.

The effect of the task content and the task format in probabilistic reasoning

Tomoko HORISHITA

Tversky and Kahnemann (1982) indicated the base rate neglect phenomenon. It implied that participants tended to neglect the base rate on the Bayesian reasoning task in which the probability information was presented in a percentage format. It is very famous as a strong and typical heuristic in probabilistic reasoning. On the other hand, Gigerenzer (1995) suggested that if change the format of the probabilistic representation from the percentage format into the frequency format, participants can solve the task better. However, in previous studies, the percentage format task required participants to answer in percentage, but the frequency format task required participants to answer in frequency. This difference suggests that the difficulty of the percentage format task depends on such task structure.

In this experiment, there are two purposes. The first is to investigate what is the difficulty of the Bayesian reasoning task. For this purpose, two kinds of question were prepared and these questions were unified between all tasks. Question 1 (Q1) was to answer in the frequency format and question 2 (Q2) was to answer in the percentage format. Question 1 emphasized the meaning of the conditional probability that is required in question 2. If the difficulty of the percentage format task depends on the task structure, the percentage of correct responses on Q2 would not be different between the percentage format task and the frequency format task. And secondly to investigate whether the advantage of the frequency format is strong or not. Two kinds of task content (risk and neutral) were prepared and compared the performance of frequency to that of percentage. If the advantage of the frequency format were strong, the manipulation of contents shouldn't affect the percentage of correct responses.

One hundred and forty-two participants were instructed to solve the Bayesian reasoning task. The design involved two probability format conditions (percentage and frequency), two question conditions (Q1 and Q2), and two content conditions (risk and neutral). Furthermore, participants were required to evaluate their imagination for each task situation.

The advantage of the frequency format was significant only in the risk content condition, and the percentage of correct responses was lower in the risk content condition than in the neutral content condition. These results mean that the advantage of the frequency format depends on the content of the task and that content of the tasks affect the percentage of correct responses. Also, these results included by the score of the imagination that showed almost the same result as the percentage of correct answers. And in all conditions, the main effect of the question was not significant and this result means that Q1 facilitated Q2 response. It suggests that the difficulty of the percentage format task depends on the task structure.