

政治的信頼の測定に関する一考察

西澤由隆*

1. はじめに

本稿は、「政治的信頼」に関する1つの測定方法について、すでに概念的には問題を指摘されながら、実証的にそれが確認されてこなかった点について、一定の検討をしようとするものである。そして、そのことをつうじて、政治的概念の測定方法に関する方法論上の、さらなる実証研究の必要性を訴えたいと考えている。

「政治的信頼」が、政治学の中心的な概念の1つであることは論を待たない。代議制のもとでは、政治的アクターや政治制度への「信頼」が、有権者の政治行動の前提として存在する。その意味では、政治に対する「信頼度」は、その国の民主主義の「成熟度」の指標と言い換えることもできる。だからこそ、世界中の先進民主主義諸国で最近観察されている「政治不信」の蔓延を、多くの政治学者や評論家が危惧をし、またその原因について盛んに議論がなされているのである。(Crozier, *et al.* [4], Norris [9], Pharr and Putnam [10])。

一連の議論での1つの大きなパズルは、「民主主義体制についての信頼感が実際に低下しているのか、それともそれは単に測定の問題なのか」との問いである。楽観的な見方をするとしたら、民主主義体制に対する根底的な支持には、それほど大きな変化はないにもかかわらず、たまたま政治学者が「的外れ」な指標を用いているために、あたかも信頼が低下しているかのように見えているだけなのかもしれない。それは単に尺度（質問の

仕方）の問題であると言うわけである。

信頼概念をどのように測定するか、その方法論上の問題はきわめて重要なインプリケーションを持つ。そして、そのことは、信頼概念に限らず、意識調査が一般的に扱うその他の概念についても、程度の差はあるとしても同様である。だからこそ、アメリカの政治学では、この方面での研究もかなりの程度進んでいる⁽¹⁾。ところが、日本の政治学の場合は、必ずしもそうではない。意識調査によるデータの蓄積という点では、この四半世紀の間にかんがりの実績を積むことができた。いよいよ、個々の質問項目の妥当性についての検証にも、リソースが向けられるべき段階に来ていると言える。

そして、そのような試みの1つとして、ここでは、「政治的信頼」尺度の妥当性についての、Hardinの指摘を取り上げることとする⁽²⁾。Hardinは、「trust」という概念は、「motivation（意図・目的、あるいは動機）」と「competence（実行能力）」の2つの要素から成り立つにもかかわらず、これまでの世論調査による実証研究は、それらを区別することがなく、信頼のメカニズムについての理解が曖昧となっていると指摘する(Hardin [5] p. 36)。信頼の二重構造についての彼の仮説が正しければ、従来の「〇〇を信頼しますか」との単純な質問では、有権者の評価のメカニズムを捉えられないことになる。そのような複合的な認知プロセスを確認するには、きめ細かな質問文が不可欠となるわけである。

このHardinの指摘を確認すべく、GLOPE 2005年調査では、「目的」と「機能」を明示的に区別した新しい指標を用意し、さらにスプリット・サンプルの手法を用いて、従来の信頼指標との比較を行った。そして、その結論を先取

* 同志社大学法学部教授

りするなら：

- ・従来の信頼指標と「目的」／「機能」の側面を明示的に区別した修正指標とでは、回答者の反応パターンは異なる。
- ・ただし、それぞれを説明変数として、他の政治的態度要因・行動要因との関連を確認したとき、必ずしも顕著な差は認められなかった。

つまり、Hardinの主張が、実証的に半ば確認できたものの、「従来の信頼指標を放棄し、今後は新指標を採用する必要がある」との強い主張を展開するには至らなかった。

以下、次のような手順で報告を進める。まず、次節では、日本の意識調査でこれまで採用されてきた「政治的信頼」についての質問文を整理する。そして、第3節では、信頼概念の二重構造についてのHardinの議論を紹介する。続いて、第4節では、実証的な検討に用いたGLOPE 2005の質問文と、比較のためのスプリット・サンプルの構造について説明し、その後、データ分析の結果を紹介する(第5節)。最後に、若干の考察と共に今後の課題を整理したい。

2. 日本における意識調査での信頼指標

日本における政治意識調査では、政治的信頼に関する指標として、現在、主に2種類の質問形式が採用されている。そして、そのうちの1組目は、かなり早い時期から意識調査には含まれてきた。たとえば、1976年のJABISS調査では、「あなたは国の政治をどれくらい信頼できるとお考えでしょうか。いつも信頼できる、大体信頼できる、時々信頼できる、全く信頼できないのうちではどれでしょうか」と聞いた上で、同じ質問を、「地方の政治」・「市区町村の政治」についても繰り返している。そして、この質問群は、JES(1986年)・JES 2(1993-96年)・JES 3(2001-05年)へと継承されている。

そして、この1組目の質問が代表制についての信頼を尋ねる限定的な質問であるのに対して、もう1組の質問形式はより一般的である。それは、より多様な政治的アクターや政治制度を個別にリスト表示し、そのそれぞれについて「信頼する／

しない」の評価を問う形式を採っている。具体的には「次にあげるものについて、あなたはどの程度、信頼することができますか」とのリード文に対して、「政党」・「政治家」・「選挙などの間接代議制」などの政治アクターや政治制度を順に提示し、それぞれについて評価を聞いていくのである。

なお、この後者の質問方式を採用している場合でも、評価の対象となる項目についての「標準的なリスト」が、日本での意識調査において一般的に定着しているわけではない。むしろ、調査により、評価対象項目の種類とその数に大きなばらつきがある。たとえば、GLOPE 2005もこの形式を採用しているが、2005年11月調査では、「テレビ」・「新聞」・「警察」から、「選挙制度」・「日本の経済体制」までの18の対象について評価を求めている(全項目リストについては、図1参照)。あるいは、JES 3(05年衆議院選挙後調査の場合)では、「NHK テレビ」をはじめとしたマスコミ各局や新聞各紙についても個別に評価を求めていることもあり、対象項目数が30に及んでいる⁽³⁾。

3. 政治的信頼のコンポーネント

さて、いずれの質問形式を採用したとしても、そのキーワードは「信頼しますか」という用語である。意識調査・世論調査の文脈で、この「信頼しますか」が、どのような意味において回答者に理解されているかを検討することは重要であろう。少なくとも、私たち研究者が想定している意味合いで、回答者がそれを理解しているかを確認する意義は大きい。

通常、システム・サポートや政治参加の研究で、政治的アクターや制度を有権者が「信頼する」という場合、それは「代理人」である政治家や政党、あるいは国が、その「主人」である市民・有権者に利益をもたらすために必要な活動を行うと、ある程度の確率で信頼することができるということだろう。たとえば、政治家を信頼するという場合は、その政治家が選挙の際に示した公約(しかもそれが投票者に利益となる公約)が、かなりの確率で実行に移されるだろうと期待できることを指すだ

ろう。また、国の政治を「信頼する」という場合、(武力によるだけでなく、経済的、あるいは文化的なものも含めた)他の国からの「侵略」に対して、自国民をしっかり守ってくれると期待ができる場合がそれに当たるだろう。あるいは、選挙制度を「信頼する」という場合は、選挙が公正に行われ、一人ひとりの票は適切に集約され、それがあらかじめ取り決められた「換算式」にのっとり、正しく議会内の議席配分として具体化されることを信じてよいということになる。

「信頼」をそのように理解すると、Hardinが指摘するように、代理人の意図・動機の適切さと、その代理人の目標達成能力の2つの評価軸の検討が内在していることがわかる。その候補者の公約は、有権者自身にとって「有利なのか不利なのか」の判断(意図に対する評価)がまずあり、それを実行に移すだけの政治力をその候補者が備えているかの判断(目標達成能力に対する評価)がそれに続く。そして、その両者の条件がそろったとき、初めてその候補者は「信頼に足る」ことになる。

これまで、信頼に関する質問項目は、「信頼しますか/しませんか」とのリード文で評価を求めてきた(ここでは、それらを「従来型」と呼ぶことにする)。ところが、もし、Hardinの指摘にしたがって、信頼のコンポーネントを考慮して測定するとしたら、「意図・目的指標」と「運用能力指標」の別々の質問を用意して、その組合せとして「信頼度」を測定する必要がある(ここでは「修正型」と呼ぶことにする)。そして、「修正型」において、回答者にとって有利な働きをすると評価され、かつ、そのように十分に機能すると判断されたとき、はじめて「信頼できる」ことになる。

もし、不利益をもたらすと考えられる評価対象(敵対的評価対象)が、その目的達成のために機能しているとしたら、それは、回答者にとってはありがたいニュースではない。「信頼できない」はずである。また、意図・目的が評価できても、それが機能していないとすれば、それは「絵に描いた餅」であるので、「信頼度」は低いかもしれない。もっとも、少なくとも目的が評価できるのであるから、一定の期待を寄せることはできるだろう。最後に、敵対的評価対象ではあるが、それがそもそも機能していないとすれば、それは

「害」にはならず、不信感を強めるほどでもない。

4. データによる検証(1)——方法

そこで、そのような判断のロジックを確認するために、GLOPE 2005 調査では、「従来型」と「修正型」に対応するような2組の質問文を用意し、それらを6種類の異なるスプリット・サンプルに対して調査を実施した。質問文とスプリットの対応を、まず確認しておこう。

4.1. 2組の信頼指標

従来型と修正型に対応する信頼指標の具体的な質問文は次のとおりである。

「従来型・信頼」指標

- ここにあげるような組織や団体については、一般的に言って、どの程度、あなたは信頼していますか。

「修正型・目的」指標

- ここにあげるような組織や団体は、その本来の目的に沿った活動をしていると思いますか。それとも、本来の目的に沿った活動をしていないと思いますか。

「修正型・機能」指標

- ここにあげるような組織や団体は、それぞれの目的を達成するためにしっかりと機能していると思いますか。それとも、機能していないと思いますか。

そして、それぞれの質問に対して、0点から10点の11ポイントの尺度を用意し、「0」から「10」の数字で回答を求めた⁽⁴⁾。それぞれ、評価の対象としたのは、図1にある18項目である。そして、同じく図1に示したように、それらを3組に別けて回答を求めた。18項目すべてを同時に提示するのは、物理的に難しいこともあるが、同時に回答者への負担感を軽減するねらいもあった⁽⁵⁾。

4.2. スプリット・サンプルの構造

異なる質問文の効果について、比較検討をしようとする場合の1つの実験デザインは、それらの質問すべてを、全員の回答者に尋ねることである。

図1 スプリット・サンプルの構造と比較パターン

評価対象	評価のタイプ	サンプル						
		A	B	C	D	E	F	
I群 (8項目) テレビ・新聞・警察・病院・小中学校・労働組合・大企業・銀行	信頼 (t)	<input type="checkbox"/>						
	目的 (i)		<input type="checkbox"/>					
	機能 (f)			<input type="checkbox"/>				
II群 (5項目) 裁判所・国会・政党・中央官庁・市区町村の役所	信頼 (t)	<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>		
	目的 (i)		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>	
	機能 (f)			<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	
III群 (5項目) 年金制度・選挙制度・日本の経済体制・国の政治・あなたの地域の政治	信頼 (t)	<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>		
	目的 (i)		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>	
	機能 (f)			<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	
	(N)	238	223	244	231	239	222	1,397

■比較可能パターン

- #1: A (I~III) vs. B (I~III) vs. C (I~III) ……信頼・目的・機能の回答パターンの比較
- #2a: D (II-t, III-t) vs. D (II-i, III-i) ……II, III評価群に対する、信頼・目的の比較
- #2b: E (II-t, III-t) vs. E (II-f, III-f) ……II, III評価群に対する、信頼・機能の比較
- #3: [F (II-i, III-i) and F (II-f, III-f)] vs. ……II, III評価群に対する、目的×機能の組合せによる新指標と従来指標の比較
D (II-t, III-t) / E (II-t, III-t)

当該の質問項目に対する回答パターンを、全サンプルについて比較検討が可能となる。最も検定力が大きいデザインである。

もともと、それが必ずしも最善の方法かという、そうとも言えない。当該の質問文間の相互の影響を排除しえない問題が残るからである。たとえば、従来型の信頼についての質問に続いて、修正型の質問が配置されたとしたら、当然のことながら、前者の后者への影響が排除できない。逆に、修正型の質問をした後に、従来型の質問をしたとしたら、この場合の「従来型」は、もはや「従来型」とは同等であるとはいえない。「目的」についての評価を経験してから、改めて「従来型」の質問をするのは、そうでない場合と状況が全く異なるからである。

さらに、今回の場合、評価対象が18項目に及ぶという量的な問題もある。回答者への負担軽減を考えると、そもそも18項目にも及ぶ質問を、3回も繰り返すことは難しい。

そこで、図1のようなスプリット・サンプルによる実験デザインを考えた。

まず、全体を無作為に6つのサンプルに別けた⁶⁾。そして、サンプルAには「従来型」指標を、サンプルB・Cに対しては、それぞれ「修正型・

目的」と「修正型・機能」を当てた。したがって、これらの3つのサンプルを比較することで、各指標の回答パターンを比較することができる(比較#1)。なお、これらの3サンプルについては、I群からIII群までのすべてを評価対象とした。

次に、サンプルのD・Eについては、いずれも、「従来型・信頼」指標を尋ねたうえで、サンプルDについては、「修正型・目的」を同じ項目に対して繰り返し尋ね、サンプルEについては、「修正型・機能」を尋ねた。なお、同一項目に対して、2回質問を繰り返すことからの負担増に配慮して、評価をお願いする項目はII群とIII群のみとした。そして、ここでの課題は、同一サンプル内で、質問形式が異なったときに、回答者の反応がどう変化するかを確認することである(比較#2)。

最後に、サンプルFであるが、評価対象II群とIII群について、目的と機能の2指標を繰り返し尋ねた。つまり、このサンプルにおいては、「目的」における評価と「機能」における評価との組合せとして、回答者の「信頼度」を測定することができることになる。そこで、このサンプルFと、「従来型・信頼」指標で測定しているサンプルD・Eとの比較において、修正モデルの優位性が確認できる(比較#3)。

図2-1 因子分析結果——因子負荷量(従来型・信頼、サンプルA)

5. データによる検証(2)——結果

分析に用いたデータセットは、GLOPE 2005 調査である。有効サンプル数は1,397。ただし、前節で紹介したように、スプリット・サンプルを用いているために、それぞれのサンプル数はおよそその1/6となっている(図1参照)。前節で紹介した3種類の比較分析の結果を順に紹介している。

5.1. 信頼・目的・機能の回答パターン比較 (サンプルA・B・C)

まず、サンプルA・B・Cを対象に、3種類の信頼指標に対して、GLOPE 2005の回答者はどのように反応したかを概括したい。

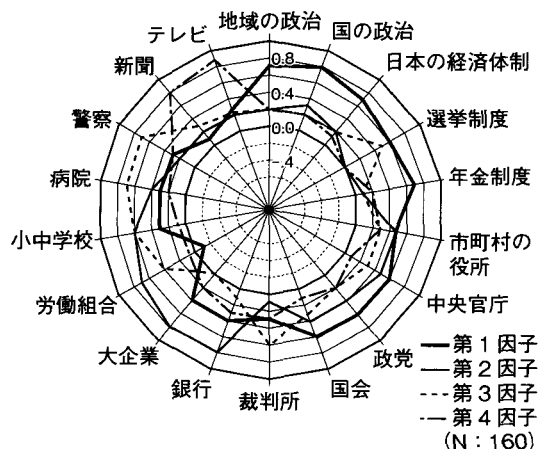
一口に「回答パターンの比較」と言っても、回答パターンの全体としての比較は容易ではない。18の対象に対する評価であり、しかも、そのそれぞれについて11ポイントの尺度による評価となっている。タテ18×ヨコ11(つまり198)のセルを持つ度数分布表を3枚用意して、それを眺めたとしても、全体像が見えそうにない。

そこで、サンプルごとに18の評価項目に対する回答を因子分析することにした。各因子に対する、評価対象ごとの因子得点の変動として、反応パターンの違いを捉えるためである⁽⁷⁾。

ここでの帰無仮説は、「質問の仕方にかかわらず、回答者は各項目に対して同じ評価点を与えた」というものである。もし、従来型指標も修正型指標も、まったく同じものを測定しているのであれば、質問文を変えたところで、抽出される因子の構造は同じとなるはずである。つまり、帰無仮説が正しければ、抽出される因子に対する各項目の因子負荷量は、どの質問を用いても変化しないはずである。

結果は、図2-1～図2-3のとおりである。

これらの図では、まず、評価の対象となる18の項目を順に放射線状に配置した。中央からの各線は、抽出された因子に対する因子負荷量を表している。中心点から周辺に向かうほぼ中間点あたりに太い円(18角形)があるが、それが因子負



荷量のプラス/マイナス0を表し、そこから外に向かってプラスの値を、中心に向かってはマイナスの値をプロットしている。もっとも外の円が因子負荷量「1」で、円の中央が「-1」に当たる。

今回の分析では、いずれの質問についても、4因子を抽出した⁽⁸⁾。図中の4つの多角形が、それぞれの因子に対応している。

まず、「従来型・信頼」指標の分析結果(サンプルA)について検討したい(図2-1)。

第1因子(太い実線)は、円の右半分には負荷量が大きく偏った配置となっていることが見て取れる。これは、項目群II(「裁判所」を除く)と項目群IIIに対応している。I群に社会・経済的アクターが含まれ、II・III群には政治的アクターが配置されていることから、第1因子は、この2グループを峻別する軸と解釈することができそうである。

第2因子(細い実線)に対しては、「大企業」・「銀行」・「労働組合」などが大きな負荷量を持っている。経済関係団体とそれ以外を峻別する因子ということになる。第3因子(破線)は、「警察」・「病院」・「小中学校」と「裁判所」について因子負荷量が大きくなっている。(裁判所についての説明が難しいが、)日常生活において関係の深い社会的アクターとそうでないアクターを峻別する因子と考えられるだろう。そして、最後の第4因子(一点破線)であるが、「テレビ」・「新聞」の2つが突出している。マスコミに関しては、その他のアクターとは別の基準で評価がされている

図2-2 因子分析結果——因子負荷量（修正型・目的、サンプルB）

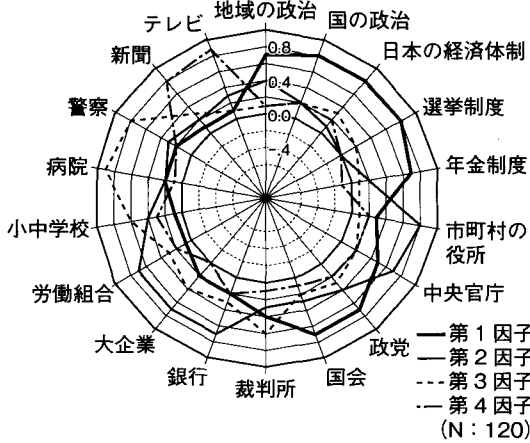
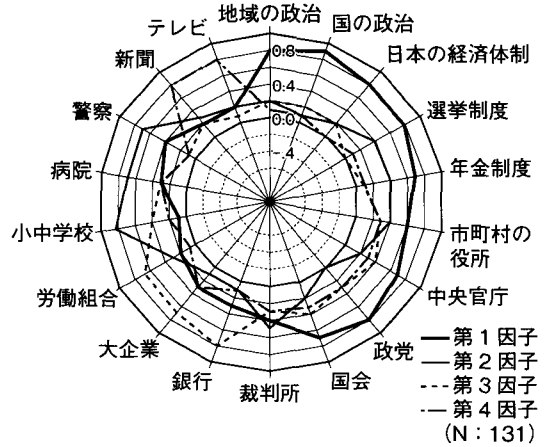


図2-3 因子分析結果——因子負荷量（修正型・機能、サンプルC）



ようである。

「修正型・目的」と「修正型・機能」についても同様に描いてみた（図2-2、図2-3）。3枚の図を上下に重ねて下から透かすように眺めると、その異同がわかりやすい。すると、「修正型・目的」についての分析結果では、第2因子において、「中央官庁」・「市区町村の役所」が経済関係団体と等しく負荷量が大きくなっている。しかし、そのことを除けば、その他の2枚のパターンとほとんど違いがないようである。そして、「修正型・機能」指標についての分析結果（図2-3）であるが、こちらでもまた他の2つとそれほど変わらない。

じつは、図2-3（機能）の第2因子、第3因子だけを個別に取り上げると、他の2図とはずいぶん様子が異なる。ところが、これは先の2図に比べて、第2因子と第3因子の順番が入れ替ったのみと見れば、結局のところは同じ構造であることに気付く。因子の抽出順は、分散の説明量の多い順に表示がされるだけで、その意味では、因子番号順が変わったとしても、因子ごとの各項目の負荷量の相対的關係が同じである限り、それは構造上の根本的な差異とはならない。

ところで、「修正型・機能」指標は、他の2つとは少し異なる点がある。それは、4つの因子の重なり程度である。「修正型・機能」指標の図は、因子ごとに見たときに、因子負荷量の大きい項目と、そうでない項目との差が、比較的是っきりとしているようである。負荷量の少ない項目については、そのほとんどが0から0.4までの範囲

で収まっているのに対して、逆に負荷量の大きいものは最低でも0.6を超え、その多くが0.8水準に達している。図において視覚的にそのことを確認するとしたら、4つの多角形のそれぞれが他と重複しない部分の面積が、図2-3において、他の2つより大きいということである。

残念ながら、そのこと（つまり、「重なりが少なすぎ」）は、統計的には確認のしようがない。しかしながら、もしその観察が正しければ、それは、それぞれのアクターが、「目的に沿って」、そしてさらに「具体的に機能しているか」と問われたときには、「信頼できるか」とか、あるいは「その目的の是非はどうか」とだけ問われる場合より、回答者には評価がしやすいと解することができる。評価の基準がより鮮明となることで、焦点のはっきりした尺度を用いて各項目を評価することになる。その結果として、評価の方向性が一致する項目と、そうでない項目の差がより明確に因子として現れてくるのではないだろうか⁹⁾。

5.2. 同一サンプル内の「信頼 vs. 目的」比較と「信頼 vs. 機能」比較（サンプルD・E）

さて、上記の比較は、異なるサンプル間の比較であった。「従来型・信頼（図2-1）」の分析対象となった回答者と「修正型・目的（図2-2）」・「修正型・機能（図2-3）」の回答者は異なる。したがって、同一の回答者が、異なる質問文に接したときに、反応パターンがどのように変化するかの確認とはなっていない。

表1 信頼度10点評価の平均値の比較——信頼 vs. 目的と信頼 vs. 機能

	対象 サンプル	信頼 (t)	目的 (i)	機能 (f)	評価差 (i - t)/(f - t)	評価差に 対する 標準誤差	t 値	危険率	(N)
裁判所	D	6.90	7.05		0.150	0.134	1.116	0.266	(187)
	E	7.00		6.90	-0.097	0.137	-0.706	0.481	(196)
国会	D	5.28	5.54		0.253	0.115	2.198	0.029*	(194)
	E	5.27		5.59	0.324	0.124	2.607	0.010*	(207)
政党	D	5.14	5.28		0.138	0.113	1.218	0.225	(196)
	E	5.20		5.37	0.166	0.115	1.443	0.151	(205)
中央官庁	D	4.69	5.19		0.494	0.110	4.477	0.000*	(180)
	E	4.88		5.22	0.340	0.111	3.074	0.002*	(191)
市区町村の役所	D	6.01	6.05		0.040	0.106	0.377	0.706	(201)
	E	5.90		6.06	0.163	0.120	1.360	0.175	(215)
年金制度	D	4.76	4.53		-0.233	0.141	-1.653	0.100	(202)
	E	5.05		5.48	0.430	0.127	3.382	0.001*	(214)
選挙制度	D	5.36	5.30		-0.059	0.123	-0.479	0.632	(204)
	E	5.53		5.81	0.279	0.115	2.437	0.016*	(215)
日本の経済体制	D	5.02	5.15		0.131	0.102	1.280	0.202	(191)
	E	5.16		5.62	0.460	0.120	3.845	0.000*	(200)
国の政治	D	4.73	5.03		0.295	0.103	2.866	0.005*	(200)
	E	4.93		5.47	0.541	0.108	5.000	0.000*	(207)
地域の政治	D	5.59	5.85		0.262	0.090	2.905	0.004*	(206)
	E	5.66		5.85	0.183	0.092	1.996	0.047*	(213)

(注) *は危険率が0.05以下のもの。

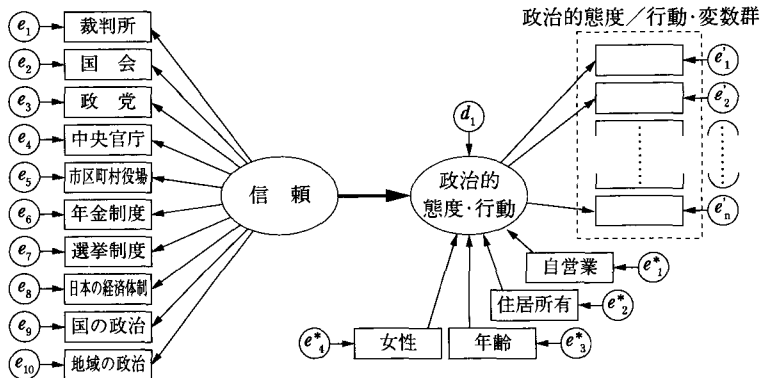
そこで、次に、サンプルD・Eを見ることにする。サンプルDは、評価対象II, IIIについて、「従来型・信頼」と「修正型・目的」の2パターンの質問に答えている。そして、サンプルEは、同様に「従来型・信頼」と「修正型・機能」の2パターンの質問に答えている。そして、もし、ここでの命題が正しければ、サンプルD・Eともに、従来型と修正型の回答パターンに差異が認められるはずである。

表1をご覧いただきたい。この表には、①II群・III群の各評価対象について、サンプルごとに、それぞれの評価項目に対する平均値を集計し、②サンプル間で、その平均値の差を求め、③その差の大きさが統計的に有意であるかの検定結果(t-テスト)が整理されている。回答者と評価項目を「対」にして、質問が2種類あるので、これは「対応のあるサンプル」に対する平均値の検定となる。「評価差」とそのt値に対する「危険率」のコラムをご覧いただきたい。仮に5%水準の危

険率で判定をするとしたら、サンプルDにおいての「(目的)-(信頼)」の「評価差」は、10項目中で4項目において有意な差が認められる。「国会」・「中央官庁」・「国の政治」・「地域の政治」については、いずれも「目的」を問われたときの方が、0.25ポイントから0.49ポイントの幅で、評価がプラスに変わっている。一方、サンプルEにおいての「(機能)-(信頼)」の「評価差」は、10項目中で7項目において有意な平均値差が認められる。サンプルDの場合の4項目に加えて、「年金制度」・「選挙制度」・「日本の経済体制」も有意となり、いずれも「機能」を問われたときの方が、0.18ポイントから0.54ポイントの幅で、評価がプラスに変わっている。質問文が変わると、反応パターンも明らかに変わる。

なお、「全項目について、同分量だけ働いている」という状況でもないことに留意していただきたい。仮にそのような分析結果であったならば、それは、質問文によって相対的な評価基準がシフ

図3 修正信頼指標——政治的態度・行動への規定力確認モデル



トしただけで、評価のプロセスが変わったかどうかは判定ができない。そうではなく、項目によって評価差にばらつきがあることからこそ、やはり、質問文は項目によっても、異なる反応を呼び起こしていると言える¹⁰⁾。

5.3. 「従来型・信頼」指標と「修正型」指標の説明力の比較 (サンプルD・E・F)

さて、最後に、新指標の優位性について検討することにします。サンプルFでは、同じくII・III群の評価対象に対して、しかも同一の回答者に、「修正型・目的」指標と「修正型・機能」指標の2つの質問がされている。したがって、このサンプルでは、両者の組合せとして「信頼度」を測定することができる。より厳密な作業定義をすることで、信頼指標の規定力・説明力が上がるかどうかを確認したいわけである。

確認の手続きであるが、図3のような一般的な因果モデルを想定し、共分散構造分析を用いて、データの当てはまりの善し悪しを、指標間で比較することにしました。図3では2つの潜在変数(楕円)が配置されているが、左側の信頼指標を説明変数として、それが影響を及ぼすだろうと考えられる、政治参加の程度・政治的忌避態度・代議制への信頼の3つを従属変数として右側に用意した。そして、後者を順次入れ替え、前者(信頼指標)の後者(政治的態度・行動)への規定力を確認した。

なお、サンプルFに対する新指標の具体的な作業定義であるが、「目的」指標と「機能」指標の積として定義した。「それぞれの指標において、評価が高いときに初めて、『真』の信頼評価とな

る」というのが、ここで検討されるべき命題であるので、そのことを具体化しようとした結果である。相互に重み付けした指標ということになる。また、サンプルD・Eについても、指標の構造がサンプルFと同じになるように、従来指標の二乗として作業定義した。そして、最後に、元の指標との値の幅(あるいは、最大値)をそろえるために、それぞれを「10」で除することにした。数式で表すと次のようになる。

$$\text{重み付け信頼指標} = \frac{\text{修正型(目的)} \times \text{修正型(機能)}}{10} \quad (\text{サンプルF})$$

$$\text{重み付け信頼指標} = \frac{\text{従来型(信頼)} \times \text{従来型(信頼)}}{10} \quad (\text{サンプルD・E})$$

ところで、ここでの信頼指標は、評価の対象がII・III群の10項目であるので、図3でも、それに対応する観測変数(長方形)が10個用意されている。一方、従属変数側の観測変数の数は、従属変数によって異なる。具体的には、「政治参加の程度」・「政治的忌避態度」については、評価の対象となる政治参加の形態は、「自治活動に積極的に関わる」から「国や地方の議員に手紙を書いたり、電話をする」までの11であるので、観測変数も11個となる。そして、「代議制への信頼」については、「国民の声を政治に反映させるのに役立つ」として「政党」・「選挙」・「国会」に評価を求める質問項目を利用しているため、対応する観測変数の数は3となる¹¹⁾。

さて、推定結果を、表2に整理した¹²⁾。期待とは異なり、「修正型」指標の説明力の優位性を確認することはできなかった。表2の下半分に、モデルの適合度を示すいくつかの指標を整理したが、CMIN・GFI (AGFI)・RMSEAのいずれの指標

表2 修正信頼指標——政治的態度／行動への規定力（共分散構造分析）

モデル サンプル	政治参加			政治的忌避態度			代議制への信頼		
	モデル1 D	モデル2 E	モデル3 F	モデル4 D	モデル5 E	モデル6 F	モデル7 D	モデル8 E	モデル9 F
■推定値									
標準化係数 (危険率)	-0.060 0.478	0.072 0.428	-0.004 0.961	0.097 0.392	-0.061 0.549	-0.030 0.787	0.183 0.038*	0.463 0.000***	0.377 0.000***
■モデルの適合度									
CMIN (確率)	685.4 0.000	755.3 0.000	728.8 0.000	591.4 0.000	631.9 0.000	569.0 0.000	411.3 0.000	435.4 0.000	368.5 0.000
GIF	0.739	0.726	0.696	0.693	0.690	0.654	0.754	0.756	0.754
AGIF	0.690	0.675	0.639	0.636	0.632	0.589	0.682	0.684	0.682
RMSEA	0.098	0.103	0.110	0.109	0.108	0.109	0.127	0.127	0.123
(N)	(157)	(167)	(139)	(99)	(112)	(92)	(156)	(168)	(141)

(注) *は危険率が0.05以下のもの, ***は0.001以下のもの。

からも、モデルのデータに対する当てはまりが良くないことがわかる⁽³⁾。さらに、信頼指標からの従属変数への係数の推計値で、統計的に有意な値を確認できたのは、「代議制への信頼」を従属変数にしたモデル7・8・9のみであった。「政治参加」や「忌避態度」については、「信頼」指標としていずれの質問形式を用いても、その規定力が確認できなかった。

標準化係数の推定値が統計的に有意であった「代議制への信頼」の3モデルについて、その係数の大小を比較してみよう。これも期待に反して、モデル9の「修正型（目的×機能）」よりも、モデル8「従来型」の規定力の方が大きくなっている。もっとも、同じく「従来型」を用いているモデル7との推定結果の違いから判断して（つまり、モデル7とモデル8とでは同じ結果が予測されるにもかかわらず、そうではないことに注目すると）、表2で推定された係数の大小からは、いかなる結論を導くことにも慎重とならざるをえない。

6. ま と め

本稿では、政治学、とりわけ政治意識研究や政治参加研究における中心的な概念の1つである「政治的信頼」について、より厳密な評価プロセ

スを具体化した作業定義の必要性を検討し、そしてそのメリットを実証的に示すことを試みた。政治的アクターの「目的・意図・動機」に対する評価と、そのアクターの「実現能力」に対する評価との合成ベクトルとして「政治的信頼」を捉え、全国規模の意識調査の中で、その優位性を確認しようとした。

分析の結果は、必ずしも「決定的」ではなかった。確かに、政治的アクターに対して異なる質問文で評価を求めた場合、回答者はそれに応じた形で異なる回答をしているようである。少なくとも、反応のパターンは変化する。そして、評価の基準・方向を明示的にする方が、回答者の意思決定が容易となっていることがデータから読み取れる。漠然とした「信頼」を問われるより、「ちゃんと機能しているか」と限定的に評価を求められた方が、答えやすいようである。

ところが、政治的信頼のメカニズムを理解しようとするときに、そのように厳密に作業定義をした方がより優位なのかということ、少なくとも今回の実験的な試みでは、その判定ができなかった。その他の政治的な態度要因や行動要因との関連性の程度を、共分散構造分析を用いて確認したが、従来の信頼指標と、より厳密に定義した修正指標とは、ほとんど差異は認められなかった。

期待した結果が得られなかったことについては、2つの解釈が可能かもしれない。1つ目は方法論

上の限界で、他方は有権者の認知処理能力上の制約である。

すでに紹介したように、今回は、6種類のスプリット・サンプルを用いた。修正指標の妥当性を検討するためには、複数の組合せでの比較が不可欠であり、そのためにはサンプルを別ける必要があった。比較パターンが増えるほど、スプリットの数も急増する。そして、スプリット数の増加は、スプリットごとのサンプル数の減少を意味する。サンプル数の減少は、検定力（パワー）の低下につながる。もしかしたら存在する「影響」が、検定力不足のために、確認できなかったかもしれない（タイプIIのエラー）。

もう1点、考慮すべきは、人の認知処理能力上の限界である。認知心理学者が繰り返し指摘するように、人は、必ずしも科学者や統計学者のように、論理的に物事を判断しているわけではない（Kelley [6], Nisbett and Ross [8], Schneider, *et al.* [11]）。信頼評価のメカニズムも、ここで検討したような論理的な判断プロセスに沿って、すべての人が、すべての状況で「厳密に行っている」のではないのかもしれない。にもかかわらず、そのように「行っている」ことを前提として、調査する側が厳密な質問文を用意したところで、一般の回答者がその「道具」を「持て余し」している可能性がある。そうだとすれば、私が想定していたような結果が得られなかったとしても不思議ではない⁴⁴。

これまで私たちは、政治学上の多くの重要な概念について、意識調査の回答者が、その用語をどのように理解しているのかという点については、必ずしも十分に吟味をせずに使ってきているように思う。そして、本稿は、そのような問題意識を出発点としている。政治的信頼の二面性の仮説については、決定的な結論には至らなかったが、世論調査における認知上のメカニズムを解明するツールとしての、スプリット・サンプル手法の有効性は示すことができたものと思っている。これからは、世論調査でのパソコンの利用が盛んとなることが予測されるが、そのことで、スプリット・サンプルの応用はもちろん、世論調査に複雑な実験的側面を持たせることが容易となるだろう。同様の問題意識に立った研究の、今後の活性化を期待したい。

[謝辞]

本稿の草稿について、早稲田大学 COE-GLOPE のワークショップ（2007年9月28日、早稲田大学）で報告をする機会を得た。その際に、討論者の久米郁男・建林正彦両氏とコーディネーターの河野勝氏から貴重な助言をいただいた。記して感謝したい。

[注]

- (1) たとえば、Sniderman [12] 参照。
- (2) じつは、「政治的信頼」尺度の妥当性についての議論は、長く続いている。たとえば、American National Election Studies (ANES) では、「How much of the time do you think you can trust the government in Washington to do what is right?」という質問が、1958年調査にすでに登場し、現在も継続して質問されている。ところが、これを含む一連の質問項目に対して、これらは、時々の個別の政治的アクターへの評価にすぎず、それが政治システム全体に対する一般的な信頼の指標となっていないとの疑問が1970年代にすでに提示されている（Miller [7], Citrin [3]）。
- (3) おそらく、この2通りの質問方式の「源流」は、アメリカの選挙研究（ANES）と、イングルハートの価値観研究（World Values Survey, WVS）のそれぞれに求めることができるのではないかと推測している。前者の質問方式は、ANESが継続的に質問を続けている。一方、後者のリスト形式は、ANESには登場せず、WVSの方で採用されている（たとえば、1981年調査）。ANESは、選挙調査に特化していることから、評価の対象を限定しているものと推測できる。一方、価値観をトータルに捉えようとしたとき、社会的なアクターをも含めて幅広く、信頼観を測定しようとするのは自然な流れであろう。
- (4) 回答者は、「回答票」と呼ばれる「選択肢カード」を手元に持ち、それを見ながら調査員の質問に答えていく。なお、そのカードには、この「尺度」が図示されており、その「尺度」の両端には、「0：信頼していない」・「10：信頼している」、「0：目的に沿わない活動」・「10：目的に沿った活動」、「0：機能していない」・「10：機能している」とラベルが付いている。また、「5」が尺度の中間点であることを明確にするために、「5」にも「中間」とラベルを付けた。
- (5) 第III群については、「修正型・目的」指標の対象の特性上、それらを別グループとして独立させる必要があった。と言うのも、第I群・第II群にある評価項目に対しては、「目的に沿った活動をしていますか」と聞くことができたとしても、「年金制度」や「選挙制度」は、それ自体が「活動をする」ようなものではないので、それらに対して「活動をしています」と聞くのは不自然であったからである。そこで、「年金制度」から「あなたの地域の政治」までの5項目は別グループ

ブとし、「それでは、ここにあげるような制度は、その本来の目的に沿うように運用されていると思いますか。それとも、本来の目的に沿わないかたちで運用されていると思いますか」と問うことにした。

(6) どのスプリット・サンプルに当たるかは、地点番号単位ではなく、個人単位でおこなった。つまり、地点ごとに、AからFの調査票をランダムに当てることとした。

(7) この表示方法は、谷口尚子氏のアイデアを参考にした。Ikeda, Ken'ich, *et al.*, "Country Report: Japan," unpublished document presented for the Asian Barometer Planning Workshop on August 11-12, 2007, Taipei.

(8) バリマックス回転後の負荷量平方和の累積が、3指標のすべてで50%を超えるように、4因子まで抽出することとした。

(9) なお、それぞれの因子分析の対象となった回答者数にも注目する必要がある。元の対象者数が、サンプルD・E・Fの順に、238・223・244であったのに(図1)、因子分析の対象となった回答者数は、160・120・131と、かなり減少している(図2の凡例の下のN)。それは、18すべての評価項目のうち、1つでも評価がないと(つまり、欠損だと)、分析から除外されるからである。分析の対象として残る「生存率」は、順に、67.2%・53.8%・53.7%であった。従来型指標が用いられているサンプルDに比べて、修正型指標の2つのサンプルでの脱落率がかなり高い。つまり、修正型指標の方が、認知負荷が高いことを示している。だとすると、因子分析に入る前段階で、回答者のスクリーニングが行われていることになる。「修正型・機能」において、「より収斂された評価が行われている」とのここでの結論についても、この「スクリーニング」の影響を否定することはできない。

(10) ただし、いずれの場合も、より好意的な方向へみの変化であることが気になる。評価のメカニズムが変わるのであれば、より好意的に評価が変わる確率と同じだけの確率で、より厳しい評価となる項目があってもよいはずである。なぜ、そのような結果となったかについては、今後の検討課題である。

(11) なお、いずれも、具体的な質問文については、次のとおり。

[政治参加・政治的忌避態度]: 「(1) この中にあるようなことについて、これまでのあなたの関わり方と、将来の関わり方について、お尋ねします。まず、「①自治会活動に積極的に関わる」ことがこれまでにありましたか。「何度かある」「1~2回ある」「1度もない」でお答えください。(2) では、そのことについて、今後はどうでしょうか。機会があれば関わりたいとお考えですか、それもとできれば関わりたくないとお考えですか」との問いにおいて、「①自治会活動に積極的に関わる」・「②役所に相談する」・「③地域のボラ

ンティア活動や住民運動に参加する」・「④デモや集会に参加する」・「⑤選挙で投票する」・「⑥選挙運動を手伝う」・「⑦候補者や政党への投票を知人に依頼する」・「⑧政治家の後援会員となる」・「⑨政党の党員となる」・「⑩政党の活動を支援する(献金・党の機関誌の購読)」・「⑪国や地方の議員に手紙を書いたり、電話をする」の11項目について回答を求めた。

[代議制への信頼]: 「ここにあげる(1)~(3)は、国民の声を政治に反映させるのに役に立っていると思いますか。それもとそうは思いませんか」との問いにおいて、「政党」・「選挙」・「国会」の3つに対して、「そう思う」から「そうは思わない」の4点尺度に沿って、順に評価を求めるものである。

なお、具体的な変数の加工手続きについては、SPSSのシンタックス・ファイルを私の「ホームページ」にて公開しているので、詳細はそちらを参照のこと。(http://ynishiza.doshisha.ac.jp/でHPを開き、メニューで“Please Download”を選択。)

(12) 表2では、紙面を節約するために、「信頼」指標の従属変数への係数に対する推定値と、モデルの当てはまりの程度を示す主な指標のみを掲載している。詳細は、同じく私のHPに、AMOSの分析結果を掲載しているので、そちらを参照していただきたい。

(13) カイ二乗値にあたるCMINについては、「モデルと観測データが合致する」が帰無仮説である。したがって、帰無仮説を棄却できないときに、「モデルと観測データが異なっているとは言えない」として、モデルがデータに適合していると判断することになる。つまり、「有意確率」が基準値(たとえば、0.05)より大きいときに「適合している」と評価することになる(クロス表などの関連性についての検定の場合とは逆の扱い)。そして、表2のモデルの場合、CMINの有意確率はいずれも基準値より小さく、モデルのデータへの当てはまりは良くない(山本他[2]37頁)。また、GFI・RMSENについては、「0.9以上」・「0.08より小さい」をそれぞれモデル採用の目安とされている(豊田[1]106頁、山本他[2]105頁)。これらの基準からも、表2のモデルはいずれも当てはまりが良くないことがわかる。

(14) 前節で紹介したように、新旧指標間で反応パターンに一定の差異があるとしても、それでも、全体として各指標はよく似たパフォーマンスを見せていた。つまり、質問文が入れ替わっても、多くの回答者が整合性のある反応をしていることになる。その意味では、もしかしたらこれは、認知能力の「限界」ではなく、不思議な認知能力の仕業と言うことができるかもしれない。異なるツールが与えられても、人は、ある一定の整合性ある反応をする能力を備えているとも考えられる。

[参考文献]

- [1] 豊田秀樹, 竹内哲監修『SASによる共分散構造分析』東京大学出版会, 1992年。
- [2] 山本嘉一郎・小野寺孝義編著『Amosによる共分散構造分析と解析事例』ナカニシヤ出版, 1999年。
- [3] Citrin, Jack, "Comment: The Political Relevance of Trust in Government," *American Political Science Review*, 68, 1974, pp.973-988.
- [4] Crozier, Michel J., Samuel P. Huntington and Joji Watanuki, *The Crisis of Democracy: Report on the Governability of Democracies to the Trilateral Commission*, New York: New York University Press, 1975.
- [5] Hardin, Russell, "The Public Trust," In Susan J. Pharr and Robert Putnam eds., *Disaffected Democracies: What's Troubling the Trilateral Countries?*, Princeton: Princeton University Press, 2000.
- [6] Kelley, H. H., "The Processes of Causal Attribution," *American Psychologist*, 28, 1973, pp.107-128.
- [7] Miller, Arthur H., "Political Issues and Trust in Government, 1964-1970," *American Political Science Review*, 68, 1974, pp.951-972.
- [8] Nisbett, Richard and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs, NJ: Prentice-Hall, 1980.
- [9] Norris, Pippa ed., *Critical Citizens: Global Support for Democratic Government*, Oxford: Oxford University Press, 1999.
- [10] Pharr, Susan J. and Robert Putnam eds., *Disaffected Democracies: What's Troubling the Trilateral Countries?*, Princeton: Princeton University Press, 2000.
- [11] Schneider, David J., Albert H. Hastorf and Phoebe C. Ellsworth, *Person Perception*, 2nd ed. Reading, MA: Addison-Wesley Publishing Company, 1979.
- [12] Sniderman, Paul M. "The New Look in Public Opinion Research," In Ada W. Finifter ed., *Political Science: The State of the Discipline II*, Washington, D. C.: American Political Science Association, 1993.