

コンジョイント実験の方法論的検討

論

宋 財 沅
善 教 将 大

説

1. はじめに

本稿の目的は、近年、政治学において人びとの選好を推定する方法として注目を集めている、Hainmueller et al. (2014) にて考案されたコンジョイント実験について、これを実施する際の指針を提示することにある。つまりコンジョイント実験を適切な形で実施するにはどのようにすればよいのかという疑問にこたえることが、本稿の課題である。コンジョイント実験は、人びとの選好を知るための有効な方法であり、今後、多くの研究者がこの方法を用いる可能性が高いと筆者らは考えている。本稿の知見は、その際のガイドラインとして活用できるものである。

近年、政治学では人びとの選好を明らかにする方法としてのコンジョイント分析に注目が集まっている。この分析手法を簡単に説明すれば、ある対象の属性 (attributes) あるいは属性の水準 (levels) が、当該対象を選択する際にどの程度考慮されるのかを統計的に明らかにする方法である。コンジョイント分析は主にマーケティングの領域で用いられているとされるが (君山 2010)、政策評価の際に用いられる場合もある (小熊・森山・横山 2015; 内田・湯沢・塚田 2015; 鷹島・桜井 2015)。さらに日本の政党選択のメカニズムの分析にこれを用いる研究もあるなど (田村・黒岩 2009, 2010)、コンジョイント分析への注目は徐々に高まりつつある。し

かし日本の政治学ではコンジョイント分析についての理解はあまり浸透しておらず、特に本稿が注目するコンジョイント実験については、その特徴や利点などが十分に理解されていない。

コンジョイント実験の特徴は、回答者に提示されるプロファイルの内容が無作為化される点にある。既存のコンジョイント分析と同様に、プロファイルを回答者に提示し、比較してもらう作業を通じて水準等の影響を推定する点はコンジョイント実験も同様である。しかしコンジョイント実験は複数のプロファイルを事前に作成し回答者に提示するのではなく、ウェブ上のプログラムによって無作為にプロファイル内容を定める。このようにプロファイルの内容を無作為化することで、より妥当な形で複数の属性や水準の効果を、統計的因果推論の枠組みに基づき推定することが可能となる。

コンジョイント実験は人びとの選好を推定するための洗練された方法であるが、その一方でこの方法には多くの課題が残存している。特に実施の際の諸問題の検討については実証研究が十分に蓄積されておらず、この問題を解決する必要があると考える。コンジョイント実験を実施する際、どの程度の観測数を確保すればよいのか。コンジョイント実験はどの程度まで繰り返して実施できるのか。satisficer 対策はどのように講じるべきか。コンジョイント実験の実施に際しては、これらを含む様々な点を考慮する必要があるが、そのための指針が現状では曖昧なのである。

そこで本稿では、コンジョイント実験の実施に際して留意すべき4つの点を明確化し、これらについての指針を、コンピュータ・シミュレーションや筆者らの実験結果などを基に提示する。具体的には、必要となる観測数、実験の繰り返し回数と脱落サンプルなどの関係性、satisficer 対策の必要性、選択肢の数の影響などを明らかにする。これらの作業を通じて、本稿はコンジョイント実験に基づく推定結果の妥当性の向上に努めると同

時に、この実験的手法を用いたさらなる実証研究の喚起に努めたい。

本稿の構成は以下の通りである。まず2.ではコンジョイント実験の概略、ならびにその意義について、コンジョイント分析や他の実験手法と比較しながら説明するとともに、残された課題について論じる。3.では実証分析に用いる意識調査の概略やコンジョイント実験案、実験の手順などについて説明する。4.では筆者らのシミュレーションや実験の結果を概観しながら、必要となる観測数や satisficer 対策などについて検討する。最後に5.で分析の結果を整理するとともに、実験実施の際のガイドラインを提示する。

2. なぜコンジョイント実験の方法論的課題に注目するのか

2.1 コンジョイント分析とは何か

政治学や行政学、公共政策学における主要な関心事の1つは人びとの選好を知ること、および選好が実際の行動などに与える影響を明らかにすることである。投票行動研究の主要な関心事の1つが政策選好の分析であることは改めて論じるまでもないが⁽¹⁾、行政学や公共政策学においても、政策需要や行政満足度を分析し把握することは重要視されている（野田2013）。代議制民主主義の下では、代理人である政治家や政党が有権者の選好に⁽²⁾応答的であることが求められる。この事実からも政策選好の分析の重要性については理解することができるだろう。

(1) 投票行動研究における政策選好の実証分析については数が膨大ということもあり、ここでの詳述は避ける。三宅（1989）、山田・飯田編（2009）、飯田・松林・大村（2015）などを参照されたい。

(2) たとえば飯田・松林・大村（2015：8-10）では、政策応答性が満たされている状態が、代議制が機能している理想的な状態だと述べられている。また政治的疎外ないし政治不信も、政府の応答性が満たされていない状態として概念化されることが多い（善教2013）。

政策選好を明らかにする方法として、もっとも一般的に用いられているのは意識調査である。郵送法や面接聴取法など実施方法の詳細については相違があるものの、いくつかの政策や争点項目を多数の回答者に提示し、それぞれについての賛否などを尋ねる点は概ねどの調査においても共通している。以下、その一例として「JESⅢ調査」の質問を見てみよう。⁽³⁾

第8回 JESⅢ調査（j波）は、2005年の郵政選挙前に実施されたものだが、そこでは集団的自衛権への賛否などについて調査がなされている。具体的には、まず「集団的自衛権の問題として、次のA、Bのような意見があります」というリード文を示した後に「A：日米安保体制を強化するためには、集団的自衛権の行使を認めるべきである、B：国際紛争に巻き込まれることになるので、集団的自衛権の行使を認めるべきではない」という対立する意見を提示する。そして「あなたの意見はどちらに近いですか」という質問文で賛否を尋ね（SQ1）、続いて「集団的自衛権の問題はあなたにとってどのくらい重要ですか」と、重要性を尋ねる（SQ2）。選択肢は、賛否が「1：Aに近い、2：どちらかといえばA、3：どちらかといえばB、4：Bに近い」であり、重要性は「1：かなり重要である、2：やや重要である、3：あまり重要ではない、4：ほとんど重要ではない」である。⁽⁴⁾

以上の従来の方法に代替する手法として近年注目を集めているのが、Hainmueller et al. (2014) にて考案されたコンジョイント実験である。この方法は一般的にマーケティングなどにて用いられているコンジョイント

(3) 正式名称は「21世紀初頭の投票行動の全国的・時系列的調査研究」である。JESⅢ調査の詳細は東京大学社会科学研究所 SSJ データアーカイブのHP (<http://csrda.iss.u-tokyo.ac.jp/> アクセス2016年2月23日) を参照のこと。

(4) わからない (DK) とこたえない (NA) も選択肢には含まれる。

分析を洗練させたものだが、日本の有権者を対象にコンジョイント分析を用いる研究はほとんどなく、そのためコンジョイント分析についての理解が現状では浸透していないように思われる。したがってコンジョイント実験を説明する前に、まず、コンジョイント分析とは何かを説明しておこう。

コンジョイント分析とは、様々な属性がもつ水準のうち、最も人びとに好まれる水準の組み合わせを統計的に分析する方法である。ここでいう属性と水準は、コンジョイント分析特有の概念であるが、選好の「構成要素」が属性、構成要素の「変数値」が水準だと理解すればよい。具体的に説明すれば、仮に筆者らが売れるラーメンを、値段、麺の太さ、麺の量、具の量、スープの種類という5点を考慮しながら開発するとしよう。これら5つの要素をコンジョイント分析では属性と呼ぶ。⁽⁵⁾ これら属性はさらに複数の水準をもつ。値段でいえば600円、650円、750円などが考えられる。これらはいかなれば値段という変数の値であるが、コンジョイント分析では水準と定義される。⁽⁶⁾

コンジョイント分析では様々な水準を組み合わせたプロファイルを複数作成し、これらを回答者に比較検討してもらう作業を通じて、各属性が選択に与える影響を推定する。再びラーメンの例を用いて説明すれば、図1に示すようなプロファイルを複数作成し、これらを回答者に提示し、好ま

(5) 何を属性とするかは分析者が任意に定める。たとえば化学調味料の有無を要素と定めたのであればそれを属性に追加すればよいし、麺の太さは考慮事項ではないと考えるならこの属性は削除すればよい。属性と水準の設定は統計分析でいうところの model specification に相当するのであり、コンジョイント分析を行う際にもっとも重視されるべきところである。

(6) 価格や重さなどは連続変数であるが、コンジョイント分析ではプロファイルを見せ比較検討してもらうという方法論上の制約からカテゴリー的な尺度として扱うことになる。そのため連続変数の場合は、どの点を閾値とするかという判断も重要となる。

しいラーメンの順に点数をつけてもらったり、もっとも好ましいラーメンを1つ選んでもらったりする。回答者に提示するプロファイルは直交表な⁽⁷⁾どを基に作成する。もっとも、属性と水準の数が多すぎると、直交表を用いても多くのプロファイルを作成しなければならない。あまりにも多くのプロファイルが回答者に提示されるような場合は推定結果の妥当性が減じられることになりかねないので、要因や水準を減らす方向で調整する必要がある(君山 2010)。もちろん、どのような属性をプロファイルに含めるのか、水準数をどのようにするのかは分析者が判断する必要がある。

	プロファイルA	プロファイルB	プロファイルC
値段	600円	650円	750円
麺	極太麺	手もみ太麺	太麺
麺の量	150g	250g	350g
具の量	マシマシ	チョモランマ	少なめ
スープの種類	豚骨醤油	塩	豚骨醤油

図1 提示するプロファイルの例

2.2 コンジョイント実験の特徴

前項で解説したのは、一般的なコンジョイント分析の方法であるが、本稿で注目するコンジョイント実験は、これとは異なる特徴をもつ。もっとも大きく異なる点は、回答者に提示するプロファイルの内容が無作為化されることである。プロファイル内容の無作為化は各水準の平均因果効果(Average Marginal Component Effect; AMCE)の推定に必要な独立性の

(7) 直交表とは、すべての水準を組み合わせた膨大な数のプロファイルを作成しなくても、比較的少数のプロファイルを作成するだけで、各属性ないし水準の効果を推定することを可能とする組み合わせを示した表のことである。

仮定を満たす際に重要となる。コンジョイント実験は、既存のコンジョイント分析以上に妥当な因果推論を行うことを可能とする。

プロフィール内容の無作為化は、コンジョイント実験の重要な特徴である。繰り返しになるが、コンジョイント実験では、回答者に提示されるプロフィール内容は完全あるいは条件付きで無作為化される。つまり調査実施者を含めて、回答者は、回答する瞬間までどのようなプロフィールが提示されるのかを知ることはできない。そのためこの方法の下では、既存のコンジョイント分析のように複数のプロフィールをすべて提示することは不可能である。逆にいえば比較的少数のプロフィールを回答者に提示し評価してもらうことを通じて、属性ないし水準の影響力を推定する。大量のプロフィールを見せる従来の方法と比較して、コンジョイント実験は回答者の認知負荷が低いという利点がある。⁽⁸⁾

コンジョイント実験は、AMCE の推定法も特徴的である。具体的には量的な従属変数だけではなく A か B かという 2 項選択型の従属変数への AMCE についても線形回帰 (OLS) を用いて推定する。⁽⁹⁾ この点についてわかりやすく説明すると AMCE は、具体的には関心のある水準以外を全て平均値に固定した上で、その水準を「存在しない状態 (0)」から「存在する状態 (1)」へ変化させた場合の従属変数 (Y) の期待値の差分として表現される。この差分はいわゆる「効果量」であり OLS では係数値に該当する。コンジョイント実験によって得られたデータの場合、条件付き

(8) 認知負荷の低さはたしかにコンジョイント実験の利点であるが、他方で認知負荷を低くすればよいというわけでもない。外的妥当性の高い推定を行うためには 1 つではなく 2 つのプロフィールを提示しなければならないなど、相対的に認知負荷が高い方法を用いなければならない場合もある点には注意されたい。

(9) 厳密には選択確率を仮定しないノンパラメトリックな手法で推定すべきだが、実際の結果は OLS でも問題ないとされる。

ロジスティック回帰分析であっても OLS であっても、各水準の効果量はほぼ一致するため、OLS で推定することに問題はないとされている。⁽¹⁰⁾

コンジョイント実験の実施方法についても簡単に解説しておこう。まずコンジョイント実験の設計をするために、Hainmueller らは Conjoint Survey Design Tool (CSDT) というツールを開発している。実験案や回答者に提示するプロファイルの画面などは、CSDT を用いて基本的には作成する。なお、このツールは Strezhnev 氏の HP から DL できる。⁽¹¹⁾ 実験案を設計後、実験を実施するための php ファイルを出力し、このファイルを、php が動作するサーバーにアップロードする。そして Qualtrics という、オンライン上でサーベイ実験を実施するためのシステムを用いれば、コンジョイント実験を実施することができる。より詳しい説明は筆者（宋）の HP を参照されたい。⁽¹²⁾

2.3 コンジョイント実験の利点

コンジョイント分析あるいはコンジョイント実験には、意識調査を用いる従来の方法と比較していくらかの利点がある。とりわけ大きな利点としては次の 2 点を指摘できる。第 1 は属性や水準の重要性について相对比较を行える点である。第 2 は過小評価バイアスを軽減できる点である。

まず、コンジョイント分析を用いる利点の 1 つとして、属性間の相対

(10) 表 3 に示すようにこの点については筆者らも確認しているが、OLS の推定結果は条件付きロジスティック回帰分析の結果と大きくは変わらない。

(11) HP の URL は <http://scholar.harvard.edu/astrezhnev/conjoint-survey-design-tool> である（2016 年 2 月 24 日アクセス）。

(12) 以下の URL にアクセスされたい。http://www.jaysong.net/studynote/methodology/qualtrics_conjoint/（2016 年 2 月 24 日時点）。なお HP のアドレスは予告なく変更する場合がある。

比較が可能だという点をあげることができる。従来の意識調査のように、複数の属性を列挙し、評定法などで重要度を測定すると、どの水準も「重要だ」と回答されることが多い。言い換えれば、ある政策争点などを「重要だ」と回答している回答者は、他の項目についても「重要だ」と回答することが多く、そのため従来の方法だと回答者の選好を曖昧な形でしか把握できないという問題が生じる。しかし、コンジョイント分析はどの属性を考慮するのかを「相対」比較の観点から検討するため、すべての属性あるいは水準について重要だと回答されるような問題は生じない。

第2のコンジョイント分析の利点は、独立変数の過小分散に基づく、過小評価バイアスを軽減できる点にある。たとえば、すべての有権者が「重要」と考え、かつ、「賛成」と考える政策に対して、仮にA政党が「反対」という立場を表明した場合、有権者はA政党を支持するか、⁽¹³⁾という問題を考えてみよう。A政党は有権者の選好とは相反する選好を表明しているので、当然、支持されないと予想される。しかし、意識調査を用いて政策選好と支持政党の関係を分析した場合、このような状況だと、高い確率で「有権者の選好とA政党支持の間には関係がない」という誤った結果が示されることになる。

なぜこのような実態に反する結果となるのか。その理由は、独立変数である重要性認識などが、先の例では無分散に近い状態だからである。従来の意識調査を用いる方法だと、「無分散デザイン」という因果推論を行う上で避けなければならない問題に（King, Keohane and Verba 1994＝2004）抵触することがある。もちろん、実際に独立変数（や従属変数）の分散が0となるデータはほとんどないが、8割以上の回答者が重要だという意

(13) 以下の記述は、有権者は自らの選好に基づき支持政党を選択するものと仮定している。

思を表明する争点項目などがないわけではない。このような場合、独立変数の分散が極端に小さくなってしまうため、当該争点の重要性が強く認識されているにもかかわらず、その変数の影響力が小さいという推定結果が示されることがある。しかしコンジョイント分析だと水準が無作為に回答者に割り当てられるため、この問題を回避することができる。

ところで平均因果効果を推定する方法としては、回答者が無作為に配分された実験などが有力な方法となる。それゆえに政治学では観察データに基づく推定法に代わり、実験的手法によって因果関係を分析する研究が、1990年代以降急増している（Druckman et al. 2011）。コンジョイント実験もその1つであるが、この方法には、他の実験的手法と比較して多くの利点がある。

たとえば他の実験的手法と比較した場合のコンジョイント実験の利点として、複数の仮説を同時に検証できる点があげられる。仮想シナリオを用いたフレーム実験の場合（秦 2015）、一般的に1つの処置群には1つの刺激のみが含まれる。もし刺激が2種類あり、さらに各刺激に2つのバリエーションが存在する場合、実験計画法に基づけば少なくとも4つの実験群が必要となる。1つの群のサンプルサイズが小さすぎると平均因果効果を十分な精度で推定できないので、結果としてフレーム実験は少数の独立変数の因果効果しか推定できない。

もちろん、分析者が1つないし少数の変数の平均因果効果のみを検証したい場合、フレーム実験は有効な手法である。しかしそうではない場合、フレーム実験の有効性は減じられる。これに対してコンジョイント実験は、複数の変数の平均因果効果を推定可能である。多数の変数の効果を推定しようとする場合、数百以上の実験群が必要とされることがあるが、コンジョイント実験はプロファイル内容が無作為化するものであり、回答者を複数の群に分割するわけではない。さらにいえば、コンジョイント実験は同

時に複数の水準の効果を推定するため、各水準の影響力の相対的な比較も可能である。

その他の利点として、コンジョイント実験は社会期待迎合バイアス (Social Desirability Bias; SDB) を除去することが可能だという点も指摘しておこう。フレーム実験では SDB を招きかねない刺激に回答者が敏感に反応することがある。つまりフレーム実験だと扱うテーマによっては刺激に効果がないのか、それとも SDB ゆえに効果が見えなかったのかを特定できないという問題が生じる。これに対してコンジョイント実験の場合、回答者がどのような属性や水準に反応したのかは秘匿されるので、人種問題などセンシティブな内容に関してであっても、安心して「ホンネ」を回答してもらうことができる。コンジョイント実験は、Randomized Response Technique やリスト実験などと同様の利点を有する⁽¹⁴⁾。

コンジョイント分析に近似する方法である Factorial Survey Experiment (FSE) との相違についても説明しておこう。FSE はコンジョイント分析と同様に、いくつかの属性よりなる「物語 (vignettes)」を回答者に提示し、それに対する評価を繰り返し行ってもらうことで、物語に含まれる属性等の効果を推定する方法である⁽¹⁵⁾。FSE はコンジョイント分析と比較して、方法論的妥当性など多くの利点があると主張されることがあるが (Auspurg and Hinz 2014)、この指摘は複数のプロファイルを同時に提示するコンジョイント分析に該当するものであり、コンジョイント実験にはあてはまらない。さらにいえば FSE よりコンジョイント実験の方が、よ

(14) Randomized Response Technique の概略については Fox (2015) を参照のこと。またリスト実験およびその投票行動研究への応用例は善教 (2016) を参照のこと。

(15) 詳しい実施方法などは紙幅の都合上、割愛する。Auspurg et al. (2015) を参照されたい。

り実態に適合的な因果推論を行える場合もある。Hainmueller et al. (2015) ではコンジョイント実験の外的妥当性が検討されているが、そこでは1つの物語を繰り返し提示するよりも、2つのプロファイルのうち好ましい方を選択してもらうコンジョイント実験の方が、より実態に適合的な推定結果となることが示されている。FSE が意義ある方法であることに相違はないが、それはFSE がコンジョイント実験以上の方法論的妥当性を兼ね備えたものであることを意味しない。

2.4 十分に検討されていない課題

コンジョイント実験には以上に見た意義が存在する一方で、多くの課題もある。特に実験実施の際の指針が曖昧であることは早急に解決すべき課題だと考える。プロファイル内容の妥当性は、分析者の目的や検証しようとする仮説によって判断すべきである。しかし実験を実施する際の観測数の問題などについては、プロファイルの内容を問わず検討することが可能である。しかしそのような研究が十分に蓄積されているとはいえない。

まず、検討されていない課題として指摘するのは先に述べた観測数についてである。コンジョイント実験は複数の属性と水準を組み合わせるため、フレーム実験などより多くの観測数が必要となる。観測数は多いに越したことはないが、費用等の制約ゆえに無限に数を増やせない。そのような場合、十分な精度で推定できる観測数の「合格ライン」があれば、実験を実施しやすくなる。しかし現状では十分にこの点について議論されていない。

実験の繰り返しの回数の上限についても曖昧である。4.2で述べるようにサンプルサイズが水準の効果を推定する際に十分ではない場合、コンジョイント実験では、同様の実験を繰り返すことによって観測数を増やす。そうすることで推定精度の高さを保つのである。しかし回答者に実験を無限に繰り返すことはできない。実質的に可能な回数は多くても10回程度で

はないだろうか。しかしこれはあくまで想定であり、実際に5回や10回という繰り返し回数を回答者が許容するかどうかは明らかではない。

第3の課題は選択肢の数、つまり測定尺度の妥当性である。コンジョイント分析の従属変数としては、AかBかといった2項選択型だけではなく、1点から10点までの点数をつけるといった順序ないしは連続変数が用いられることも多い。どのような選択肢が望ましいかは難しい問題であるが、比較的はっきりとした態度を表明しにくいとされる日本人の回答傾向を考慮した場合（田崎・二ノ宮 2013）、4件順序のような曖昧な選択肢を含む方が望ましいという発想はあり得るだろう。いずれにせよ選択肢の数は、記述的推論や因果推論の相違を生じさせる原因となることが実証されており（善教 2015）、これがいかなる結果の相違をもたらすかは、検討する価値があるといえる。

第4はsatisficerに関する問題である。筆者らが利用するオンライン調査に限らず、意識調査には常に自己の「効用」を最大化しようとするsatisficerが存在する（三浦・小林 2015a）。satisficerの存在は、時に推定結果を歪める原因ともなるので（三浦・小林 2015b）、コンジョイント実験の実施に際しても、この問題については検討しなければならない。satisficerはどのようにしたら識別可能なのか。またsatisfice傾向を抑止するためにはどのような対策が有効か。さらにsatisficerを分析結果から除外することで推定精度は高まるのか。これらの疑問についても検討すべきだろう。

以上より本稿ではここまでの議論で述べてきた4つの残された課題を中心に、コンジョイント実験を実施する際の指針を検討する。なお本稿は、コンジョイント実験を行う際の参照基準となる実践的ガイドラインを提示することを目的とするが、ここで示す知見は、あくまで筆者らの実験デザインと結果に基づくものだという点で限界を抱える。本稿の示す基準はあ

くまで「参照」程度であり、唯一の解とすべきではない。しかし基準が曖昧な現状においては、たとえそのような限界を抱えるものであっても意義ある知見だと考える。

3. 実験のデザイン

3.1 調査の概略

本稿では、筆者らが2015年12月11日から12月15日にかけて実施した意識調査を用いた分析結果に基づき、前節で述べた課題を検討する。この意識調査は、コンジョイント実験の方法論的検討を目的に実施したものであり、設問数は最大で10問、うち8問がコンジョイント実験に関わる質問である。残る2問のうち1問は年齢に関する質問であり、もう1問は政治的アクターなどへの感情温度である（自民党、民主党、安倍晋三、橋下徹）。有効回答者の総数（N）は2000人である。⁽¹⁶⁾

コンジョイント実験は Qualtrics を用いて実施した。また回答者は、Yahoo! クラウドソーシングにタスクを依頼する形で収集した。Yahoo! クラウドソーシングの回答者は、全国の有権者から無作為抽出されたサンプルではないため代表性に劣る。本稿の目的はあくまでコンジョイント実験の方法論的検討にあり、全国有権者の政策選好を推定することにはないので、サンプルの偏りが大きな問題となるわけではない。しかしどのような属性をもつ回答者なのかを明示しておく必要はあるだろう。筆者らが実施した調査には、コンジョイント実験とは関係のない質問がほとんど含まれていないが、年齢と政党への感情については調査している。これら2つの変数の分布を以下では確認する。

まず年齢の最小値は18、最大値は84と、変数値の幅は広い。しかし平

(16) 途中 Drop まで含めると2157人である。

均値は約39, 標準偏差は約9.2であり, サンプルの多くが30代から40代である。実際に割合を見ても, それぞれ10代が0.8%, 20-24歳が3.4%, 25-29歳が9.3%, 30-34歳が17.9%, 35-39歳が19.6%, 40-44歳が21.7%, 45-49歳が14.2%, 50-54歳が7.4%, 55-59歳が3.2%, 60歳以上が2.4%となっている。Yahoo! クラウドソーシングから得られた回答者には, 若年層と高齢者が少ないという偏りが存在する。

次に感情温度について見ると平均値(標準偏差, 有効回答者数)は, それぞれ自民党が45.3 (22.2, 1798), 民主党が32.8 (21.4, 1797), 橋下徹が46.9 (25.7, 1807), 安倍晋三が45.1 (25.8, 1821)であった。一般傾向との相違は, 同様の意識調査を全国の有権者を対象に同時期に実施していないのでわからない。しかし2014年12月の衆院選前後に筆者(善教)は全国の有権者を対象に意識調査を実施しており, この結果と比較することで, 本稿の調査の回答者がどの程度偏ったサンプルなのかを確認できる⁽¹⁷⁾。なお, この調査は楽天リサーチに登録しているモニタから2635人を, 性別, 年齢, 居住都道府県が全国の有権者の分布と近くなるように抽出した上で, 様々な政治意識などを尋ねたものである。そこでの感情温度の平均値は, 自民党が46.63 (27.0, 2372), 民主党が31.25 (22.8, 2355), 橋下徹が42.6 (27.2, 2399), 安倍晋三が48.61 (28.0, 2405)であった。本稿で用いるデータと大差なく, 党派性のバイアスは小さいと考えてよいだろう⁽¹⁸⁾。

(17) この調査は「有権者の政治的無知・政治能力と政治文化の比較研究」(代表者: 山田真裕) プロジェクトの一環として実施されたものであり, 筆者(善教)は分担者の1人である。

(18) 本稿のデータでは, 自民党への感情温度がやや低めに見積もられている印象を受ける。しかしそれは高齢層が少ないデータであることに起因する可能性が高く, Yahoo! クラウドソーシング特有の問題というわけではない。

3.2 コンジョイント実験案

筆者らの調査では、2つのパターンの実験を実施している。その概略は表1に整理する通りである。なお、本稿の目的はあくまでコンジョイント実験の方法論的検討にあるので、この実験案の内容についての説明は省略する。

実験1は「今から、2人の社会問題の解決を試みる候補者が擁立され

表1 無作為化コンジョイント実験の属性と水準

実験1 *属性の順序は固定	
属性	水準
所属政党	自民党 / 民主党 *この水準は固定であり無作為配分されない
議員定数	現状維持 / 1割減 / 2割減
重視する利益	国よりも地元の利益 / 地元よりも国の利益
雇用政策	転職市場の活性化 / 新卒採用枠の拡大 / 退職年齢の引き上げ
福祉政策	全般的な社会保障の縮小 / 出産・育児手当の拡充 / 高齢者医療費保証の拡充/失業手当の拡充
年金対策	給付金開始年齢の引き上げ / 納付額の引き上げ / 給付額の引き下げ
消費税	8% (現状維持) / 5% (引き下げ) / 10% (引き上げ)
支持率	候補者A優勢 / 伯仲 / 候補者B優勢

実験2 *属性、水準すべてランダムイズ	
属性	水準
市区町村議会の定数	現状維持 / 現状から1割削減/現状から2割削減 / 現状から3割削減
自治体職員数	現状維持 / 現状から1割削減/現状から2割削減 / 現状から3割削減
まちづくりの主体	市区町村がまちづくりを推進/都道府県がまちづくりを推進 / 国がまちづくりを推進
都市計画	まちの中心部を集中的に開発 / 中心部と周辺部を同程度開発 / まちの周辺部を集中的に開発
医療や福祉	市区町村単位でサービス供給 / 都道府県レベルでサービス供給 / 身近な行政区を設置し、そこでサービス供給
教育政策	市区町村単位でサービス供給 / 都道府県レベルでサービス供給 / 身近な行政区を設置し、そこでサービス供給
図書館運営	市区町村と都道府県の図書館を両方維持 / 市区町村と都道府県のどちらかに統廃合
公営住宅	市区町村と都道府県の公営住宅を両方維持 / 市区町村と都道府県のどちらかに統廃合

注) 実験1の「支持率」はプロファイル間でランダムイズしていないため、分析の際は除外している点に注意されたい。

た総選挙で、投票してもらいます」というリード文を示した後に、「仮に、次回の総選挙で次のような公約を掲げた2人の候補者が立候補していると想定してください。あなたは、どちらの候補者に投票したいと思いますか」という質問文で、2つの異なるプロフィールのうち、棄権を含め、どちらかのプロフィール（候補者）を選択する、というものである。

実験2は「続いて、地域の問題の解決を試みる2人の候補者について、どちらを支持するのか、回答して頂きます」というリード文を示した後に、「仮に、次の政策を掲げた2人の候補者が立候補していると想定してください。あなたは、どちらの候補者を支持しますか。支持する候補者を1人だけ選んでください」という質問文で、2つの異なるプロフィールのうち、どちらかのプロフィール（候補者）を選択する、というものである。なおここでは選択肢の異なる2つのパターンをさらに用意しているが、これは次節の分析と関連するため、次項にて詳しくその理由を説明する。

3.3 課題の検討方法

まず、観測数の検討方法から説明する。必要観測数については、本調査のデータを用いるのではなくシミュレーションから検討する。コンジョイント実験で必要となる観測数に関しては、これまでいくつかの議論がある。その中の1つである「関心のある水準が500回以上出現」（Orme 2010）という基準を発展させた「すべての水準が500回以上出現」という基準を満たす「回答者×繰り返し回数」を、シミュレーションを用いて分析する。

次に、繰り返し回数について検討する方法を説明する。実験の繰り返しはsatisficerの増加を帰結するのか、それとも回答者の「回答慣れ（効率化）」を帰結するのか。また調査脱落者との関係はどうか。これらの点を明らかにするために本稿では実験1について、繰り返し回数が異なる3つの群（2回、3回、4回）を作成した。そして回答者をこの3群

のうち1つに無作為に配分する。そして群間の推定結果の比較、および回答時間の比較分析を行い、複数回繰り返すことの問題を分析する。

選択肢の違いの効果を検証する方法としては、実験2について、2項選択型と4件順序型の2群を用意し、回答者をこの2つのいずれかに無作為に配分する。そして2項選択群と4件順序群の推定結果を比較する。具体的には4件順序の値を2項選択に再割り当てし、その推定結果が2項選択群の推定結果に近似するかを確認する。

最後に satisficer に関しては、本稿の調査だけでなく、筆者らが行った別調査のデータを補助的に用いて検討する。なぜなら本稿の調査には satisficer を識別するための質問が含まれていないからである。したがって、satisficer の識別質問が設けられている別の調査から回答者の回答時間を確認し、satisficer を識別するための回答時間の「閾値」を推定する。その結果を本調査データに適用できるかを、本研究では検討するのである。satisficer の識別方法などは次節で説明する。

4. 分析結果

4.1 コンジョイント実験結果の概略

実験実施に係る諸課題の検討を行う前に、コンジョイント実験の結果を提示しながら、分析結果の解釈方法を説明する。コンジョイント実験でどのようなことを明らかにできるのかを示すためである。ただし紙幅の都合上、ここで解説するのは実験1の結果のみとする。

図2は、前節で示した実験1の推定結果を整理したものである。実験1の選択肢には「棄権する」もあるが、これを含めて結果を解釈すると複雑になるため、棄権を選択した回答者は欠損として⁽¹⁹⁾いる。つまり図2は

(19) 棄権の選択人数(割合)はタスク1では280名(13.28%)、タスク2
84(628) 法と政治 67巻2号 (2016年8月)

「棄権する」を選択した回答者を除いた状態で、政策選好をコンジョイント実験によって推定した結果ということになる。

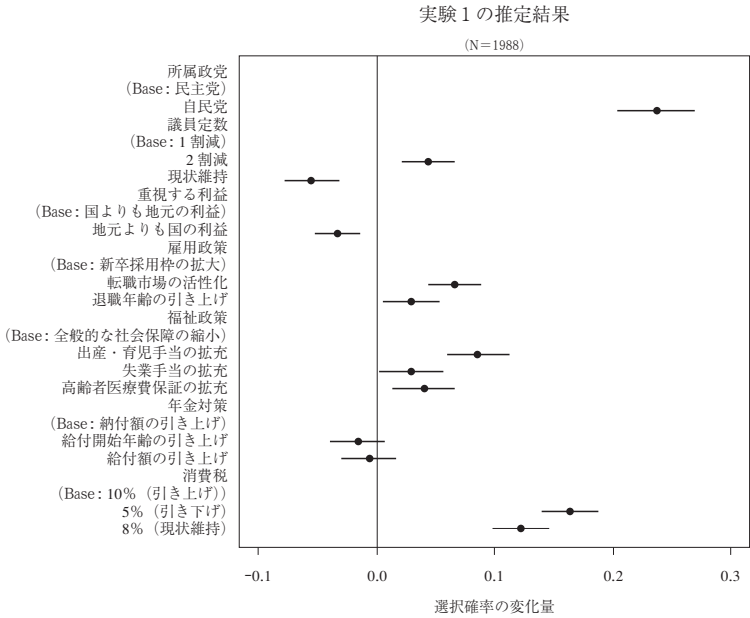


図 2 コンジョイント実験の結果 (実験 1)

さて、この図 2 は、回帰プロットやキャタピラプロットなどと呼ばれるものであり、OLS の結果を図示化したものである。黒丸は左側に示されている水準の効果の推定値であり、横に伸びる棒は推定値の 95% 信頼区間である。たとえば第 1 の所属政党の効果について説明すると、効果の参照基準 (ベースライン) は民主党であり、これを基準とする場合の自

では 256 名 (12.19%)、タスク 3 では 79 名 (11.30%)、タスク 4 では 80 名 (11.48%) である。

民党の係数は約0.23である。これは候補者が民主党の場合と比較して自民党だと、回答者から選択される確率が約23ポイント増加することを意味する。つまり所属政党への選好は回答者の選択確率を約23ポイント変化させる効果をもつと解釈できる。

SDBの除去という点からも図2を検討しておこう。実験1にはどの利益を重視するかという属性も設けている。推定結果は、全体の利益よりも地元の利益を重視する候補者の方が回答者に好まれるというものであった。一般的に利益誘導政治に対しては批判が多いが、回答者の「ホンネ」としては、国の利益を重視する候補者よりも地元の利益を重視する候補者を望む人が多いという結果である。コンジョイント実験はSDBの影響を受けにくいことがこの結果からわかる。

以上にくわえてコンジョイント実験は、どの属性の影響が大きいのかという相対比較も容易である。たとえば年金給付額の引き上げを見ると、どの水準の効果も有意ではないことから、これはそれほど回答者に重視されていない属性だと解釈できる。他方で消費税については、現状維持も引き下げも、ともに10ポイント以上選択確率を左右する。くわえて消費税の影響は社会保障や雇用などよりも大きい。ここから、相対的には（所属政党を除いて）消費税に関する政策が、もっとも意思決定に影響を与え⁽²⁰⁾と考えることができる。

さらにコンジョイント実験は、推定結果を用いて「もっとも好まれるパッケージ」が何かも予測できる。これは逆に「もっとも好まれないパッケージ」が何かも予測可能であることを意味する。属性ないし水準の違いの総合的な効果の大きさを計算可能であると同時に、ある政策パッケージがもっとも好まれる、あるいは好まれない点からどの程度離れているのかを分析

(20) もちろん以上は実験案に妥当性があることを前提とするものである。

できることもコンジョイント実験の大きな利点である。

表2は、図2の結果に基づき、好まれる水準で構成された政策パッケージとそうではない水準で構成されるパッケージの差分を計算したものである。差分の総和は約0.68であり、これはもっとも好まれないパッケージと好まれるパッケージの選択確率の差が約68ポイントあることを意味する。この表で示しているのは最も好まれるパッケージともっとも好まれないパッケージであるが、当然、様々な政策パッケージ案を比較することが可能である。

表2 仮想候補者の政策内容（括弧内は推定値）

	もっとも好まれる 組み合わせ	もっとも好まれない 組み合わせ	差分
所属政党	自民党(0.236)	民主党(0.000)	0.236
議員定数	2割減(0.044)	現状維持(-0.056)	0.100
重視する利益	国よりも地元(0.000)	地元よりも国(-0.033)	0.033
消費税	5%(0.164)	10%(0.000)	0.164
福祉政策	出産・育児手当の拡充(0.085)	一般的な社会保障の縮小(0.000)	0.085
雇用政策	転職市場の活性化(0.065)	新卒採用枠の拡大(0.000)	0.065
計			0.683

なお実験1の従属変数は2項選択型であるが、平均因果効果の推定方法としてはOLSを用いている。コンジョイント実験では、ダミー変数の従属変数に対する独立変数の効果の推定方法がOLSであってもよい。このことは2.2にて既に述べた通りであるが、OLSと条件付ロジットの推定結果を比較し、改めて近似する結果となるのかを確認しておこう。

表3は実験1データを用いた条件付ロジットによる推定結果と、OLSによる推定結果を比較したものである。両者の結果がほとんど変わらないことがわかるだろう。所属政党属性内の自民党を除く水準の係数の差分は全体的に1ポイント前後であり、大きな差があるとはいえない。表3は、水準の効果の推定方法は条件付きロジットでもOLSでもどちらでもよい

ことを意味するが、OLS による推定には条件付きロジットと比較して、推定結果を直感的に解釈しやすいという利点がある。また条件付きロジット推定においては、ある独立変数の限界効果が他の変数の値によって変化する一方で、OLS の係数の推定値は互いに独立している⁽²¹⁾ので、ある水準の効果を解釈する際に他の水準を考慮する必要がないことも利点として指摘できる。

表3 条件付きロジットとの比較

	条件付きロジット		OLS
	推定値(SE)	効果量	推定値(SE)
所属政党			
自民党	0.811(0.000)***	0.158	0.236(0.016)***
議員定数			
2割減	0.191(0.000)***	0.037	0.044(0.011)***
現状維持	-0.250(0.000)***	-0.050	-0.056(0.011)***
重視する利益			
地元より国の利益	-0.141(0.000)***	-0.028	-0.033(0.010)***
雇用政策			
転職市場の活性化	0.288(0.000)***	0.055	0.065(0.011)***
退職年齢の引き上げ	0.123(0.012)*	0.024	0.029(0.012)*
福祉政策			
出産・育児手当の拡充	0.378(0.000)***	0.071	0.085(0.014)***
失業手当の拡充	0.134(0.018)*	0.026	0.029(0.013)*
高齢者医療費保証の拡充	0.171(0.003)**	0.033	0.040(0.014)**
年金対策			
給付開始年齢の引き上げ	-0.067(0.171)	-0.013	-0.016(0.012)
給付額の引き下げ	-0.022(0.654)	-0.004	-0.007(0.011)
消費税			
5% (引き下げ)	0.722(0.000)***	0.132	0.164(0.012)***
8% (現状維持)	0.532(0.000)***	0.100	0.122(0.011)***
観測数	11030		11030

注) *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

(21) ただし交差項を投入していないことを前提とする。

4.2 分析① 必要となる観測数

コンジョイント実験は、大量の観測数を推定の際に必要とする。ここでいう観測数とは、回答者数、繰り返し回数、プロファイル数の積である。すなわち観測数は回答者数×繰り返し回数×プロファイル数となるので、たとえば1000人の回答者に2項選択型の実験を3回繰り返すと6000（1000×3×2）となる。6000という観測数は従来の意識調査から見れば十分大きいですが、コンジョイント実験では、各水準が必要十分な回数、回答者に提示される必要がある。したがって水準数が極端に多い場合は膨大な観測数が必要になることもある。

コンジョイント実験は提示するプロファイル数や属性の水準数によって十分な観測数が異なるが、他方でそのことは、観測数の基準を設けることができないことを意味しない。たとえばその基準として、 $\frac{nta}{c} > 500$ が考案されている（Johnson and Orme 2003）。ここで n は回答者数、 t は繰り返し回数、 a はプロファイル数であり、 c は最も多い水準を含む属性の水準数である。仮に繰り返し回数が4、最大水準数が8の2項選択実験の場合、この指針に基づけば回答者は500人以上いれば十分ということになる。また Orme（2010）は、関心のある水準が500回以上出現する必要があるというさらなる基準を提示している。ただしこの500回は必要最低限の基準だともされている。

コンジョイント実験の長所の1つである複数の仮説の同時検証を可能とするには、関心のある水準ではなくすべての水準が500回以上出現する必要がある。したがって本稿では、Orme（2010）の基準を発展させ、「すべての水準が500回以上出現する」という新たな基準を提案する。では、観測数がどの程度であればこの基準は満たされるのか。本稿ではシミュレーションからこの問題を検討する。

シミュレーションの概略を述べる。まず、仮想的なコンジョイント実験

案を設定する。回答者に提示するプロファイル数は2としている。観測数について検討する際、もっとも重要なのは最大水準数であるが、ここでは「最大水準数4」と「最大水準数8」の2パターンとする。⁽²²⁾そして実験を複数回、仮想状況下で行い、その中でもっとも出現数が少なかった水準の出現回数が「安定的」に500回と750回を超える観測数を計算する。詳しいシミュレーションの設定は筆者（宋）のHPを参照されたい。⁽²³⁾

図3はシミュレーションの結果を整理したものである。図3の左上の図を見ると、プロファイル数が2であり、最大水準数が4、最低出現回数を500と設定した場合に必要な「回答者数×繰り返し回数」は約1500であることがわかる。Orme (2010) の基準では1000あれば十分となるが、筆者らのシミュレーションの結果に基づけばこれは十分ではない。さらに最低出現を750回に設定した場合を見ると（右上図）、必要とされる「回答者数×繰り返し回数」は2200程度まで増加する。安定した結果を得るためには、既存の基準の約2倍の観測数が必要である。

最大水準数が8の場合は、さらに多くの「回答者数×繰り返し回数」が必要である。筆者らのシミュレーション結果によると、最低出現回数を500回にしても、最大水準数が8だと「回答者数×繰り返し回数」は約2600必要である。実験の繰り返し回数が4回であれば650人の回答者が必要だということである。また最低出現回数の基準を750にした場合、「回

(22) 水準の数が4となる実験案は多いが、8を超えることは少ないだろう。いわば水準数4は「典型例」、水準数8は「最大値」として理解すればよい。

(23) シミュレーションのコードは宋のHPに公開している（URL: http://www.jaysong.net/research/cjoint_n_sim/ 2016年2月24日）。なおコンジョイント実験は、かなり低い確率ではあるが、両プロファイルに同じ属性ないし水準が出現する場合がある。その場合は、カウントしない設定でシミュレーションを行っている。

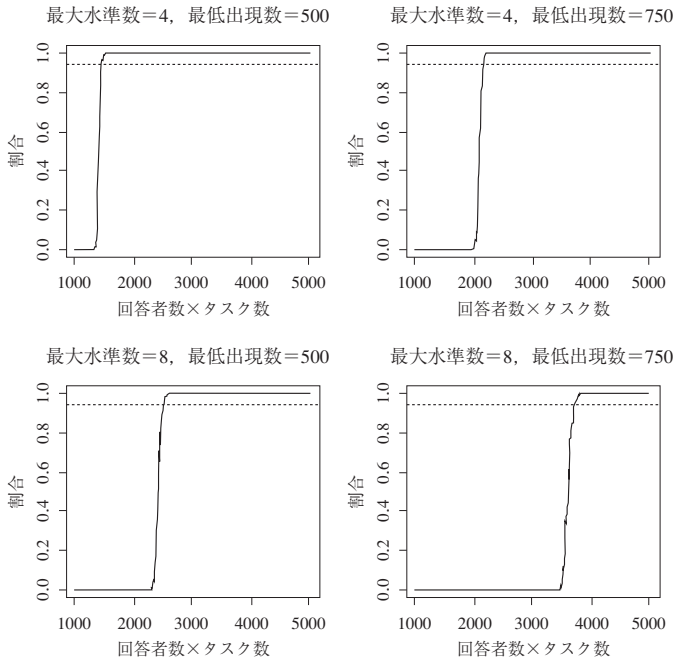


図3 必要観測数のシミュレーション結果

「回答者数×繰り返し回数」は約3800必要である。

以上のシミュレーションの結果をもとに、観測数のガイドラインを述べるならば次の通りとなろう。プロファイル数が2のコンジョイント実験を行う際、最大水準数が8以内であれば「回答者数×繰り返し回数」は2600程度あればよい。具体的にはプロファイル数が2の実験において、回答者が1000人いた場合、3回実験を繰り返せば十分な精度で水準の効果を推定できる。回答者が500人の場合は5回もしくは6回、実験を繰り返す必要がある。もっともこれは欠損サンプルが0、またサンプルを分割しないという状況での基準である。Satisficerの除外など、サンプルサイズが何らかの形で小さくなることを想定するなら、繰り返し回数は下限

となる回数から1回程度増やしておいた方がよい。

4.3 分析② 繰り返し（タスク）数

コンジョイント実験は、前項で述べたように十分なサンプルサイズが確保できていれば、実験を繰り返す必要はない。他方、十分な数の回答者を確保できなかった場合は実験を繰り返す必要がある。しかし、あまりにも多く実験を繰り返すと、回答者が調査から脱落する可能性がある。脱落しないとしても不適切な回答をする回答者が増え、そのことが推定結果を歪める可能性がある。したがってこの繰り返しの回数についても、実験実施の際は検討する必要がある。

まずは繰り返し回数と効果の大きさの関係から確認していこう。3.3で述べたように、実験1では、回答者は、繰り返し回数が2回から4回までの群に無作為に割り当てられている。2回繰り返し群はグループ1（657人）、3回はグループ2（656人）、4回はグループ3（668人）とする。図4はこのグループごとに各水準の効果の推定値をプロットしたものである。図中の線は水準の効果の推定値の推移であり、繰り返し回数が増えると効果の推定値がどのように変化するかを整理したものだと理解すればよい。

全体的な傾向を確認すると、繰り返し回数と効果の推定値の変動はほとんど関係がない。したがって実験を繰り返すことで推定結果が歪む傾向はないと考えられる。また、キャリーオーバー効果もほとんどないといつてよい。ただし一部の水準の推定値、とりわけ属性1の水準1、属性4の水準3は、実験回数が増えるにしたがって効果の推定値が大きくなる傾向にある。くわえて数は少ないが、実験ごとに推定値の符号の向きが異なるものもある。

図4に見られる一部の効果の不安定性は、サンプルサイズの小ささに
92(636) 法と政治 67巻2号 (2016年8月)

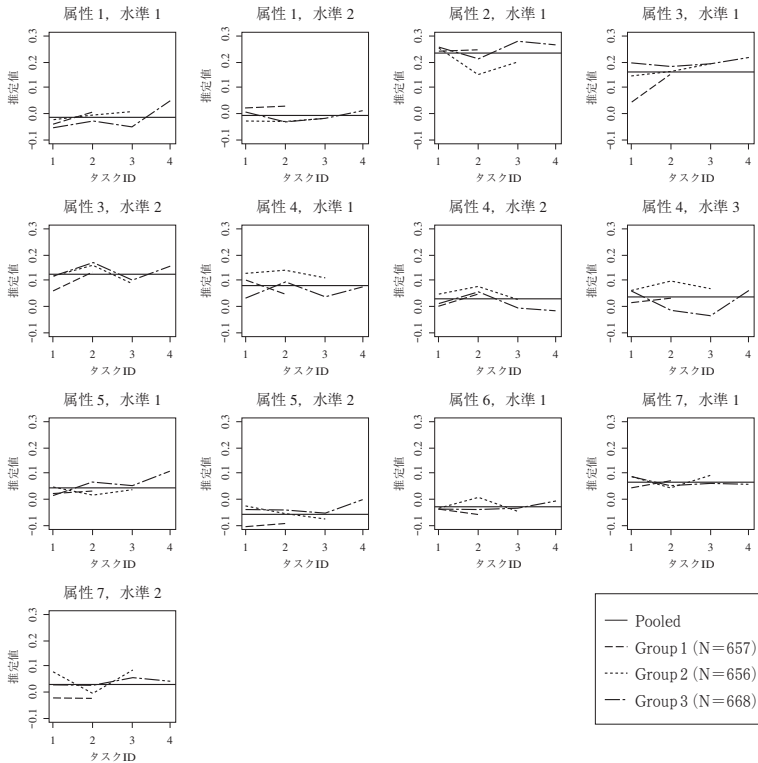


図4 繰り返しによる推定値の変化

起因すると考えられる。先に述べたようにシミュレーションの結果から筆者らは、最大水準数が4の場合「回答者数×繰り返し回数」は約1500必要であると述べた。図4の推定結果のサンプルサイズはこの基準の半分未満であり、ゆえに推定値の信頼性は低下せざるを得ない。このことが推定値の不安定性をもたらす原因となっている。⁽²⁴⁾

(24) なお、図4のように視覚的に表現した場合、変動の幅だけを見て判断するのは危険である。変動幅が0.1であっても、元の推定値が0.01だったのか、それとも0.5だったのかによって解釈は異なる。どの程度の変動に

続いて、繰り返し数と回答者の Drop 率の関係を分析する。コンジョイント実験は、類似の実験を複数回繰り返す。複数の属性からなるプロファイルと比較することにくわえて、その作業をさらに複数回繰り返すのだから、実験の後半になるほど回答を途中で諦める誘因が生じると考えられる。表 4 は、実験 2（二項選択型）の結果を整理したものであるが、この表を見ながら、繰り返し回数と「回答疲れ」の関係を確認する。

表 4 タスク数と脱落者割合

タスク	脱落(人)	累積脱落(人)	脱落率(%)	累積脱落率(%)	残り(人)
1	0	0	0.00	0.00	1041
2	3	3	0.29	0.29	1038
3	6	9	0.58	0.86	1032
4	7	16	0.68	1.53	1025

表 4 を見ると、たしかに後半になると脱落者が増える傾向を確認できる。しかし全体的にその割合は小さく 2 % にも満たない。実験 2 の前には実験 1 を行っていることを想起すれば、回答者は 6 回から 8 回のコンジョイント実験に回答したことになる。それでも調査から脱落する回答者は多くない。傾向としては実験後半になるほど Drop 率は高くなるが、繰り返し回数の増加が Drop 率に与える影響は小さい。

では、実験の繰り返し回数が増えるたびに、回答者は不適切な回答をするようになるのだろうか。表 5 は、タスクごとの平均回答時間を整理したもののだが、たしかに後半のタスクになるほど、平均回答時間が短くなっている。ただし実験 1 と実験 2 の平均回答時間はほとんど変わらない。実験 2の方が実験 1 より属性数が多いため、実験 1 よりも考慮する時間

意味があるかは分析者の判断に委ねられるが、僅かな変動をもって安定していないとみなすことにも慎重になるべきであろう。

が長くなることを筆者らは想定していたが、実態としては両者の間には差がない。

表5 タスクごとの平均回答時間（秒）

タスク	実験1	実験2（2項）	実験2（4件）
1	35.8	36.4	38.8
2	20.6	21.3	23.7
3	17.1	18.5	18.8
4	15.6	16.6	16.4

注）画面が表示されてから選択肢を最後にクリックするまでの秒数をカウントした

ここで問題となるのは、表5に示す傾向（平均回答時間の低下）をどのように解釈するかである。図4で示したように繰り返し回数の増加にともなう係数の歪みなどはほとんどなく、また脱落者も少ないことから、回答者は実験が繰り返されるにつれて、徐々に提示される属性や水準について学習していると考えられる⁽²⁵⁾。しかしそれは satisficer の増加を意味するわけではなく、どのような属性が出現するのか、またどのような水準の違いに注目するのかなどを学習しているために、2回目以降の回答時間が短いと考えた方がよいだろう。

以上の結果をまとめよう。実験の繰り返し回数を多くすることは、少なくとも5回程度までは問題はない。繰り返し回数の増加による回答者の脱落は少ない。回答者は実験が繰り返されるにしたがって慣れていくが、それは回答者が satisficer になることを意味しない。ただし10回、20回といった、かなり多い繰り返し回数については、何らかの問題が生じる可能性はある。FSE では繰り返し回数の上限は10回程度とされている

(25) これは適応から生じるキャリーオーバー効果を指すのではなく、単にコンジョイント実験の形式に慣れていくことを意味する。キャリーオーバー効果が生じる可能性も否定できないが、図4の分析結果からはそのような影響があるとはいえない。

(Auspurg et al. 2015)。回答者への負担を考慮するなら、コンジョイント実験でも10回程度を繰り返しの上限とすべきであろう。

4.4 分析③ 選択肢の数

選択肢の数はどのように設定すればよいのだろうか。2項選択型にした方がよいのか、それとも曖昧な選択肢も含める4件順序型にすべきなのか。筆者らが実施した実験2には、「A/B」の2項選択型と4件順序型の2つのパターンを用意した。どちらのパターンに回答者が配分されるかは無作為であるため、推定結果の差は選択肢の効果ということになる。なお選択肢のコーディングについては、2項選択型は選択した方を1、されなかった方を0とし、4件順序選択型は選択肢を2値に再割り当てした上で（AとどちらかといえばAを統合）、2項選択型と同様にコーディングした。

選択肢の数の違いの効果を分析した結果を整理したものが表6である。2項選択型も4件順序型も、全体的な傾向はほとんど変わらない。もちろん推定値の有意確率が異なるなど、細かな点での相違は見られる。しかしいずれも効果の差は誤差の範囲を超えるものではない。選択肢の違いが劇的な結果の差を生じさせるとはいえない結果である。

ただし全体の傾向として、2項選択型の方が4件順序型よりも効果の推定値が大きい傾向は確認できる。統計的に有意な差とはいえないが、数でいえば16の水準のうちの13の水準の効果について、2項選択型の方が大きな値である。どちらの結果が真に正しいのかを判断することは現時点ではできないが、選択肢の違いが推定結果の違いをもたらす可能性は否定できない。

なぜ4件順序型にすると推定値が小さくなるのかは判然としない上に、真値が不明である以上、2項選択型と4件順序型のどちらが妥当かを述

表6 2項選択群と4件順序群の比較

	(2項選択) 推定値 (SE)	(4件→2件) 推定値 (SE)	論 説
まちづくりの主体 (Base: 国)			
市区町村がまちづくりを推進	0.109 (0.014) ***	0.107 (0.014) ***	説
都道府県がまちづくりを推進	0.076 (0.014) ***	0.066 (0.014) ***	
公営住宅			
市区町村と都道府県の公営住宅を両方維持	-0.031 (0.011) **	-0.028 (0.011) *	
医療や福祉 (Base: 市町村単位)			
身近な行政区を設置し、そこでサービス供給	0.018 (0.013)	0.006 (0.013)	
都道府県レベルでサービス供給	-0.031 (0.014) *	-0.026 (0.014)	
図書館運営 (Base: 統廃合)			
市区町村と都道府県の図書館を両方維持	0.021 (0.011)	-0.002 (0.012)	
市区町村議会の定数 (Base: 1 割削減)			
現状から2割削減	0.067 (0.016) ***	0.040 (0.015) **	
現状から3割削減	0.099 (0.016) ***	0.073 (0.015) ***	
現状維持	-0.054 (0.015) ***	-0.069 (0.015) ***	
教育政策 (Base: 市町村単位)			
身近な行政区を設置し、そこで教育内容を決定	0.027 (0.014) *	0.009 (0.014)	
都道府県レベルで教育内容を決定	0.025 (0.014)	0.014 (0.014)	
自治体職員数 (Base: 1 割削減)			
現状から2割削減	0.022 (0.015)	0.048 (0.016) **	
現状から3割削減	0.054 (0.016) ***	0.050 (0.016) **	
現状維持	-0.062 (0.016) ***	-0.051 (0.015) ***	
都市計画 (Base: まちの中心部)			
まちの周辺部を集中的に開発	0.006 (0.013)	0.021 (0.014)	
中心部と周辺部を同程度開発	0.038 (0.014) **	0.037 (0.014) **	
観測数	8296	8296	

注) *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$

べることは難しい。この疑問にこたえるには Hainmueller et al. (2015) のような、推定結果の外的妥当性を検証するさらなる分析が必要であろう。ただ、曖昧な選択肢を設けることのバイアスの問題や (善教 2015)、いたずらに選択肢を増やすべきではないといった点を考慮するなら、4件順序型よりは2項選択型の方が、現時点では望ましいように思われる。⁽²⁶⁾

4.5 分析④ satisficer の識別と対策

コンジョイント実験は、回答者の認知負荷が従来の意識調査よりは高い方法である。さらに実験を繰り返すことをある程度念頭におき設計することが通例である以上、適切ではない回答をする satisficer にどのような対策を講じるかは重要な関心事となる。satisficer の識別法は既にいくつか考案されているが、本稿ではそのうち回答時間に注目する。つまり回答時間が短く、プロフィールを比較していないと考えられる回答者を satisficer とみなす。

回答時間に注目する際、問題となるのは satisficer と非 satisficer を識別する「閾値」をどこに設定するかである。3 秒や 5 秒、あるいは 10 秒といった基準を恣意的に設けるのも 1 つの方法だが、本稿では、別の意識調査を用いた分析から「閾値」を推定し、それを本稿の調査結果にあてはめつつ、satisficer と非 satisficer を区別する意味について検討する。

筆者らは、本稿とは別の研究プロジェクトとして、大阪市民・府民を対象とするオンライン意識調査を 2015 年 12 月に実施した。より詳しく述べれば、大阪在住の有権者（20 歳～79 歳男女）2205 人、大阪市を除く、大阪府在住有権者（20 歳～79 歳男女）2590 人に意識調査を実施しており、調査項目の中には、satisficer を識別するための質問、およびコンジョイント実験に関する質問を設けている。satisficer はマトリクス型質問の中に「この項目は「ややそう思う」を選択してください」を設け、ややそう思う以外を選択した回答者を satisficer と識別した。satisficer の割合は全体の約 25%（1212 人）であった。

この大阪のデータを用いて satisficer か否か（satisficer=1, 非

(26) 2 項選択と 4 件順序の比較だけではなく、6 件順序や 8 件順序とも比較する必要があるが、本稿では検討できていない。今後の課題である。

satisficer=0) を従属変数、コンジョイント実験（繰り返し 2 回）の回答時間を独立変数とするロジット推定を行ったところ、回答時間と satisficer の間には統計的に有意な関係があることが明らかとなった。その結果を整理したものが表 7⁽²⁷⁾である。回答時間が短くなるほど、satisficer である確率が高くなるという推定結果であり、これは直観的にも納得できるものである。

表 7 satisficer を従属変数とするロジット推定の結果

	係数値	誤差	Cox-Snell R ²	的中率
回答時間（タスク 1・大阪府）	-1.110	0.059	0.158	78.8%
回答時間（タスク 2・大阪府）	-1.141	0.062	0.153	78.3%
回答時間（タスク 1・大阪市）	-1.007	0.061	0.149	78.8%
回答時間（タスク 2・大阪市）	-1.128	0.063	0.166	78.8%

注）係数値はすべて $p \leq 0.001$ で統計的に有意

この推定結果に基づき、satisficer として識別される予測確率が 50% を超える回答時間を計算したところ、1 回目の実験における回答時間では 6.11 秒、2 回目では 3.99 秒という結果となった。なお予測確率が 60% を超える閾値は、1 回目だと 4.08 秒、2 回目だと 2.79 秒であり、70% を超える閾値は 1 回目だと 2.64 秒、2 回目だと 1.88 秒であった。本稿の調査における実験内容と大阪府民を対象とする実験内容は異なるものであるが、平均回答時間は酷似している。そのためこの satisficer 識別のための「閾値」を本稿のデータにも適用できる余地はあるといえる。

回答時間によって回答者を除外すると、効果の推定値と標準誤差がどのように変わっていくのかを確認しよう。実験 1 のタスク 1 と 2 のデータに含まれる回答者を、最後にクリックした秒数を基準に除去した上で係数

(27) 推定ミスを未然に防ぐため、回答時間が極端に長い回答者（300 秒以上）は欠損とした。ただしその数は 10 名程度であり少ない。

と標準誤差を推定した結果を整理したものが図5である。各線は1つの水準である。ここでは効果と標準誤差の変動に関心があるため、各線が具体的にどの水準を表すのかは省略する。横に伸びる複数の線があるため解釈しづらいかもしれないが、横軸が0の場合は回答者を除去しない場合の推定値であり、10の場合は最後のクリックが10秒以下の回答者を除去した際の推定値であるので、左から右に回答者を除外する基準が変化していくにつれて、効果の推定値や誤差がどのように変動しているのかを示すものだと理解すればよい。

図5を見ると、効果の推定値も標準誤差も、5秒前後までは、回答者を削除しても値がほとんど変わらないことがわかる。しかしそれ以降の秒数の回答者を除去した場合、大きな変動が生じている。回答者を除去するタイミングの下限値が大きくなるにしたがって、効果の推定値は不安定さを増し、標準誤差の値も線形的に増加している。この5秒という値は、⁽²⁸⁾

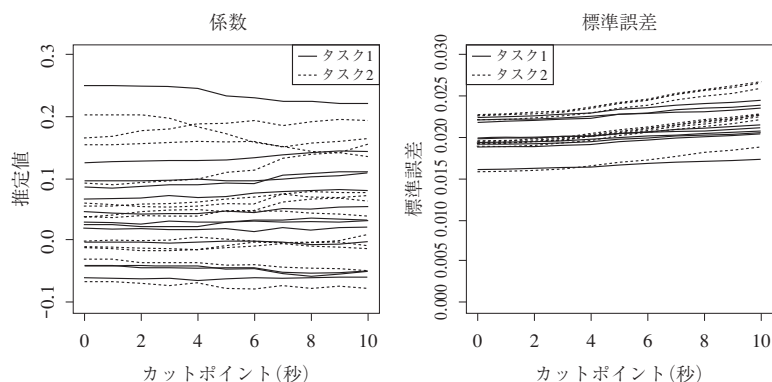


図5 回答時間と係数（推定値）・標準誤差の関係

(28) 標準誤差は観測数の平方根の逆数と比例するため、サンプル脱落による標準誤差の拡大も含まれる。

大阪府民を対象とするデータの分析結果から得られた「閾値」とほぼ一致していることは明らかである。つまり効果や誤差の変動が5秒前後まで見られなかったのは、除外された回答者の多くが satisficer であったからだと考えられる。

satisficer の存在が効果の推定値や標準誤差に影響を与えるのであれば、これを識別する必要性が生じる。しかし図5に示したように、satisficer の影響力は大きくない。つまり、コンジョイント実験では、satisficer を識別する必要性はそれほど高くないことを以上の結果は示唆する。satisficer による推定値の歪みはほとんど生じないし、誤差も大きくなるわけではない。ただし satisficer が増えることで推定結果の信頼性が低下してしまう可能性はある。全回答者が satisficer であれば、コンジョイント実験の推定結果には意味がなくなるし、また先に検討したシミュレーションの基準も無意味となるので、回答者の satisficer 化を未然に防ぐ方法を考える必要はある。

ここでは、その方法の1つとして、回答時間の制約を設ける方法について検討する。先に述べた通り、satisficer を識別する1つの目安は、1回目の実験では6秒程度、2回目以降では4秒であった。これを基準に考えると1回目の実験の回答時間を最低5～6秒、2回目以降は最低3～4秒経過しないと、次の調査画面に進めないという対策を講じることが、satisficer 化を未然に防ぐ1つの方法として有効ではないだろうか。

5. 結 論

5.1 分析結果のまとめ

本稿では、コンジョイント実験を実施する際の課題について実証的に検討した。具体的には、実験実施に際して必要となる観測数はどの程度か、実験はどの程度まで繰り返して実施してよいのか、選択肢はどのようにす

べきか、satisficer 対策はいかにすべきかという 4 点について、シミュレーションや筆者らの実験結果に基づき検討した。

まずは第 1 の点である、必要観測数の分析結果をまとめておこう。必要観測数は水準の最大数によって変化することを念頭におきつつ、以下ではシミュレーションによる分析結果に基づき、最大水準数を 4 とする場合と 8 とする場合の 2 パターンについて述べる。提示するプロファイル数を 2 とし、かつ最大水準数を 4 とする場合、最大水準の出現数を 500 程度確保したいならば「回答者×繰り返し回数」は約 1500、750 程度確保するなら約 2100 必要である。また最大水準数を 8 とする場合、「回答者×繰り返し回数」は約 2600、750 程度であれば約 3800 である。提示されるプロファイルの数と水準の最大数が「回答者×繰り返し回数」を考える上では重要となるが、最大水準数が 8 を超えることはそれほど多くないように思われる。したがって、プロファイル数が 2 の場合、全体で 4000 程度の「回答者×繰り返し回数」を確保できていればよい。

続いて実験の繰り返し回数について検討する。繰り返しの回数の限界がどこかという問題は、必要な観測数はどの程度かという問題と関連する。実験を繰り返し実施する理由は、観測数を増やす点にあるからである。回答者の認知負荷の観点からいえば繰り返し回数は少ない方がよい。つまり、推定の際に必要な観測数を満たすにはどの程度繰り返せばよいのかを考えるべきであり、複数回繰り返すことを前提に実験計画を設計する必要はない。プロファイル数が 2 の場合、回答者数が 1000 人であれば 4 回繰り返せばよいし (4×1000)、2000 人なら 2 回でよい (2×2000)。回答者数が十分に確保できる場合、必要以上に実験を繰り返す必要はない。

繰り返し回数の上限は、大規模サンプルに基づくサーベイ実験ではなく、学生など多くの回答者を用意できない場合に問題となる。プロファイル数が 2 で回答者総数が 400 人の場合、4000 という本稿の基準を満たすには 10

回程度、実験を繰り返す必要がある。その際、10回も繰り返すことに問題は無いのかを検討しなければならない。筆者らは、(異なる実験の組み合わせだが)最大で8回、実験を繰り返す設計のもとで、コンジョイント実験を実施した。その結果、8回程度であれば実験を繰り返しても問題は無いことが明らかとなった。筆者らの調査結果では、実験を複数回繰り返しても途中脱落者が増加しない⁽²⁹⁾。繰り返し回数が10回未満であれば問題は無いと判断してよいのではないだろうか。

選択肢の数について検討した結果も整理しておこう。結論を先取りすれば、本稿の実験結果は、AかBかという2項選択型と、A、どちらかといえばAという4件順序型の間に大きな相違はないというものであった。一部、水準の効果の有意確率に違いは見られたが、おおよその傾向は変わらず、さらに部分的な違いも誤差の範囲を超えるものではない。現実社会の意思決定に近づけるなら2項選択型の方がよいように思われるが、曖昧な選択肢を設けるべきという意見もあるだろう。この点については、もう少し検討する余地があるが、現状では2項選択でも推定結果に歪みなどが生じるわけではないことを指摘しておきたい。

コンジョイント実験は、多数の属性ないし水準を比較してもらう点で一般的な意識調査よりも認知負荷が高く、そこから satisficer が増える可能性があると考えられる。本稿ではこの問題についても検討し、効果の推定値や標準誤差は satisficer を除外しても変化しないことを明らかにした。satisficer がデータに含まれるからといって、推定結果の妥当性や信頼性

(29) FSE では繰り返し回数は10回程度までとされているが、コンジョイント実験では30回程度まで繰り返しても大きな問題は無いという結果もある。ただ、たとえ妥当な推論が行えるとしても、回答者への認知負荷の観点から、30回も同じ質問を繰り返すことは避けた方がよいのではないかと考える。途中脱落されるリスクを念頭におけば、同じ実験のタスク数は10回を限度に設定した方がよいのではないだろうか。

が著しく低下するわけではない。

しかし、回答者全員が satisficer であれば、当然、推定結果は大きく歪む。このような極論はさておくとしても、多くの回答者に適切な回答をしてもらうために、satisficer 化を阻止する対策を講じることは必要だろう。本稿ではその 1 つとして、回答に制限時間を設けるという方法を提案した。その際、どの程度の制限時間とするかが問題となるが、1 回目では「5 秒」、2 回目以降は「3 秒」を下限として設定すればよいだろう。これは大阪府民を対象とする実験から得られた基準であるが、本稿のデータにも適用可能であった点に鑑みれば、一般化可能な基準の可能性はある。あまりにも長い秒数を制約として課すことは避けるべきだが、3 秒から 5 秒程度の制約であれば問題はないように思われる。

5.2 今後に向けて

コンジョイント実験は、政策選好と意思決定の関係を明らかにする方法として、今後、大いに活用されていく可能性がある。コンジョイント実験はより望ましい政策の組み合わせは何かを検討することも可能な方法であり、さらに既存のコンジョイント分析や他の実験と比較しても利点を多く有する。本稿で明らかにした知見は、そのようなコンジョイント実験を実施する際の指針として活用可能である。

もちろん、本稿はすべての疑問や課題を検討するものではない。たとえば、実施の際に問題となる使用デバイスの問題については十分に検討できていない。オンラインモニターに登録している回答者のうち、いくらかはスマートフォンを用いて回答している。PC とスマートフォンでは調査画面が異なり、これが因果効果の推定に何らかの影響を及ぼす可能性がある。また、本稿は方法論的検討を行うことを主眼としているため、実験案の作成方法についても議論できていない。属性と水準の設計が不適切なもので

あれば、当然、推定結果の妥当性も担保されないことは、改めて指摘するまでもない。

さらに、本稿で用いた意識調査のサンプルは、全国の有権者から無作為に抽出されたものではないので、どこまで一般化できるのかという疑問もある。選好の推定結果を除いて、必要観測数や繰り返し回数など、本稿で検討している問題の大部分はサンプルの特性にそれほど依存しないと考えている。ゆえに本稿の知見や指針は一般化できる可能性があると考えているが、より代表性の高いサンプルを用いて、同様の分析を行うべきであろう。さらにシミュレーションの方法についても、もう少しパターンを増やすなどする必要もある。

謝辞：

本稿は2015年12月に神戸大学で開催された神戸大学政治学研究会における報告「コンジョイント分析の方法論的検討」に、修正と加筆を行ったものである。研究会参加者の方々、ならびに本研究に対して有益なコメントを提供してくださった三輪洋文氏と坂本治也氏に、ここに記して感謝申し上げる次第である。

付記：

本稿は、科学研究費助成金若手研究B（課題番号15K16995）、および科学研究費助成金基盤研究B（課題番号26285036）の研究成果の一部である。

参 考 文 献

Auspurg, K. and T. Hinz (2014) *Factorial Survey Experiments*. SAGE Publications, Inc.

———, T. Hinz, C. Sauer, and Stefan Liebig (2015) “The Factorial Survey as

- a Method for Measuring Sensitive Issues,” in Uwe Engel, et al. *Improving Survey Methods: Lessons from Recent Research*. Routledge, pp. 137-149.
- Druckman, James N., Donald P. Green, James H. Kuklinski, and Arthur Lupia (2011) “Experimentation in Political Science,” James N. Druckman et al. *Cambridge Handbook of Experimental Political Science*. Cambridge University Press, pp. 3-12.
- Fox, James A. (2015) *Randomized Response and Related Methods: Surveying Sensitive Data*. SAGE Publications, Inc.
- Hainmueller, J., Daniel J. Hopkins and Teppei Yamamoto (2014) “Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments.” *Political Analysis*, Vol. 22, No. 1, pp. 1-30.
- , D. Hangartner and Teppei Yamamoto (2015) “Validating Vignette and Conjoint Survey Experiments against Real-world Behavior.” *Proceedings of the National Academy of Sciences*, Vol. 112, no. 8, pp. 2395-2400.
- 秦正樹 (2015) 「非有権者における政治関心の形成メカニズム：政治的社会化の再検討を通じて」日本選挙学会2015年度大会報告論文。
- 飯田健・松林哲也・大村華子 (2015) 『政治行動論』有斐閣。
- Johnson, Rich and Bryan K. Orme (2003) “Getting the Most from CBC.” *Sawtooth Software Research Technical Paper*.
- 君山由良 (2010) 『第3版 コンジョイント分析』データ分析研究所。
- King, G., R. O. Keohane and S. Verba (1994) *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press (=真淵勝監訳 [2004] 『社会科学のリサーチ・デザイン：定性的研究における科学的推論』勁草書房)。
- 三浦麻子・小林哲郎 (2015a) 「オンライン調査モニタの Satisfice に関する実験的研究」『社会心理学研究』31巻1号, pp. 1-12。
- 三浦麻子・小林哲郎 (2015b) 「オンライン調査モニタの Satisfice はいかに実証的知見を毀損するか」『社会心理学研究』31巻2号, pp. 120-127。
- 三宅一郎 (1989) 『投票行動』東京大学出版会。
- 野田遊 (2013) 『市民満足度の研究』日本評論社。
- 小熊仁・森山治・横山壽一 (2015) 「過疎地域における乗合バスサービスの住民評価と政策的課題：北陸鉄道奥能登バス三崎線に対するコンジョイント分析」『運輸と経済』2015年8月号, pp. 57-71。
- Orme, Bryan K. (2010) *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*. Second Edition, Research Publishers

LLC.

鷹島充寿・桜井慎一（2015）「市民が望む調整池の親水公園化整備に関する研究：常時滞水型調整池に対するアンケート調査およびコンジョイント分析結果」『日本建築学会計画系論文集』Vol. 80, No. 711, pp. 1139-1146.

田村征洋・黒岩祥太（2009）「コンジョイント分析による有権者の政策選択に関する研究」『日本オペレーションズ・リサーチ』Vol. 52, pp. 1-19。

田村征洋・黒岩祥太（2010）「コンジョイント分析による有権者の政党選択に関する研究」『行動計量学』Vol. 37, No. 1, pp. 105-117。

田崎勝也・二ノ宮卓也（2013）「日本人のレスポンス・スタイル：構造方程式モデリングを用いた探索的研究」『社会心理学研究』29巻2号，pp. 75-85。

内田倫彦・湯沢昭・塚田伸也（2015）「表明選好法による都市基幹公園の防災機能の便益評価に関する検討」『都市計画論文集』Vol. 50, No. 3, pp. 409-415。

山田真裕・飯田健編（2009）『投票行動研究のフロンティア』おうふう。

善教将大（2013）『日本における政治への信頼と不信』木鐸社。

———（2015）「日本の社会は「不信社会」か？：サーベイ実験による政治的信頼指標の妥当性の検証」『法と政治』66巻1号，pp. 109-136。

———（2016）「社会期待迎合バイアスと投票参加：リスト実験による過大推計バイアス軽減の試み」『法と政治』66巻4号，pp. 1-26。

補遺 提示するプロフィールの例（実験1）

	候補者A	候補者B
*所属政党	自民党	民主党
*議員定数	2割減	2割減
*重視する利益	国よりも地元の利益	国よりも地元の利益
*雇用政策	転職市場の活性化	新卒採用枠の拡大
*福祉政策	全般的な社会保障の縮小	失業手当の拡充
*年金対策	給付額の引き上げ	納付額の引き上げ
*消費税	5%（引き下げ）	8%（現状維持）
*支持率	候補者A優勢	候補者A優勢

あなたは、どちらの候補者に投票したいと思いますか。

候補者A
（赤色）
○

候補者B
（青色）
○

棄権する
○

On Conjoint Survey Experiment: A Methodological Investigation

Jaehyun SONG
Masahiro ZENKYO

This study aims to examine the methodological subjects in a conjoint survey experiment. One of the most important topics in political science is to clarify what preferences citizens have and how the relationship between citizens' preferences and political behavior. While several previous studies have used the survey method as the most conventional method to estimate preferences, it has some limitations and shortcomings. The conjoint survey experiment, proposed by Hainmueller et al. (2014) as an alternative method, is a sophisticated method to estimate preferences in citizens, because it is able to reduce the social desirability bias and estimate preferences more appropriately based on the framework of causal inference in statistics. Therefore, recent studies have focused on the conjoint survey experiments. Moreover, studies that estimate citizens' preferences based on this experimental method have been gradually increasing in political science.

However, Japanese political scientists have almost never used the conjoint survey experiment; therefore, many methodological points are obscure. In particular, the following four points should be given due consideration when we conduct a conjoint survey experiment: 1) the necessary number of the observations, 2) the relationship between the repetition of experiments and dropouts from experiment, 3) the valid number of choices, and 4) the measures for satisficer. Through the analysis of these four points based on our original survey data, this study proposes and suggests a useful guideline for conducting the conjoint survey experiment.