

## オンライン調査・実験におけるsatisfice：その影響と対処方法

小林哲郎（香港城市大学）・三浦麻子（関西学院大学・大阪大学）

連絡先（小林哲郎）：[tkobayas@cityu.edu.hk](mailto:tkobayas@cityu.edu.hk)

オンライン調査・実験は実証政治学におけるデータ収集の手法の1つとしてすでに定着した。国内外の雑誌や学会で発表される研究の多くがオンライン上で実施された調査や実験に基づいていることは周知の事実である。オンライン調査は、確率標本抽出が困難であるという欠点が初期の頃から指摘されてきたが、ネット利用人口が拡大するにつれてその簡便さと低廉さという魅力が相対的に増大し、厳密な代表性を欠くという留保付きで広く用いられるようになった。また、オンラインでは無作為配置を伴う実験操作も容易に行えるため、調査研究のみならず実験研究も多く行われている。当初はオンライン広告などを通して集められた自己選択型のパネルを構築したオンライン調査会社を通して調査や実験を実施することが一般的であったが、近年ではAmazon Mechanical Turk (AMT)などのオンラインクラウドソーシングプラットフォームを通して研究者が直接的に参加者をリクルートすることが一般的になっている。

オンラインでのデータ収集は確かに簡便かつコストが低く、学生サンプルを用いた実験室実験などと比較してより多様な参加者をリクルートすることが可能であるため、メリットが大きい。しかし、一方ではオンラインで収集されるデータの質について十分な検討がされているとは限らない。本稿では特に、オンライン調査・実験において参加者が目的を達成するために必要最小限を満たす手順を決定し、追求する行動（以下、行動をsatisfice、こうした行動を取る回答者をsatisficerとする）に注目する。オンライン調査・実験における参加者の目的とは、多くの場合研究への参加を完了して謝礼を受け取ることである。その目的を達成するために必要なコストを最小限化するために、たとえば論理的に不正解とならない中間回答（「どちらともいえない」や「わからない」など）を選択することで教示文や調査項目を読み飛ばしたり、ランダムな回答を行ったり

することがありえる。こうした十分な認知的リソースを割かない回答行動は調査や実験の結果を歪める危険性がある。

データ取得過程におけるsatisficeはオンライン調査・実験以前から知られていた現象であるが、オンラインでのデータ収集では特に重要な問題になる。まず、オンライン調査・実験では研究者が回答行動を直接監督することができないため、十分な注意を払わない回答行動が生じやすい。質問項目や実験刺激への注意を促す教示文を入れたとしても、ブラウザの別のタブでソーシャルメディアを見たり、あるいはテレビを見ながら回答するなどの行動が発生することを排除できない。さらに、ランダムサンプリングに基づいた調査とは異なり、回答者は自己選択的に調査や実験に参加する。しかし、一回の調査や実験への参加で得られる報酬は少額のため、まとまった額の報酬を得るために多くの調査・実験に参加するというインセンティブが発生する。すなわち、オンライン調査・実験の参加者は、多くの研究参加を効率的に「こなす」ことになりがちである。こうしたオンライン調査・実験の特徴はsatisficeの発生を促進すると考えられる。本稿では著者らが行ってきた一連のsatisfice研究をまとめ、それが国内外でどの程度広く生じているのか、satisficeはどのように研究の知見を歪めるのか、satisficeに対してどのような対策を取ることが有効なのかという3点について論じる。

## Satisficeの発生状況

オンラインでのデータ収集においてsatisficeはどの程度一般的に生じているのだろうか。三浦・小林(2015a)は、日本国内のオンライン調査会社2社のパネルを対象に、Oppenheimer et al. (2009)のInstrumental Manipulation Check (IMC)を用いたsatisficeの検出を試みた。オリジナルIMC設問は図1の通りである。教示を読み飛ばして選択肢の上の部分だけ読むと、普段行っているスポーツを尋ねる質問のように読める。しかし、実際には教示文の最後に、「画面上部のタイトルをクリックするように」との指示がかかれており、これによって教示文を読んでいるかどうかという注意のレベルが測定できるようになっている。

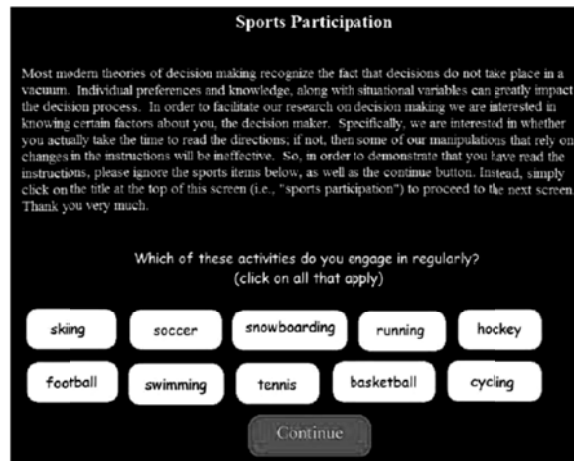


図1 オリジナルのIMC設問 (Oppenheimer et al. (2009) のFig. 1)

三浦・小林(2015a)はオリジナルのIMCを改変し、図2に示すものを用いた。これによって回答者の注意のレベルを事前に測定した上で、後続のリッカート尺度の測定において「この項目は一番左(右)を選択してください」といった項目に対して指示通りに回答するかどうかを観察した。なお、その際にリッカート尺度の項目数について10項目、30項目、50項目の3水準を設け、実験参加者を無作為配置した。

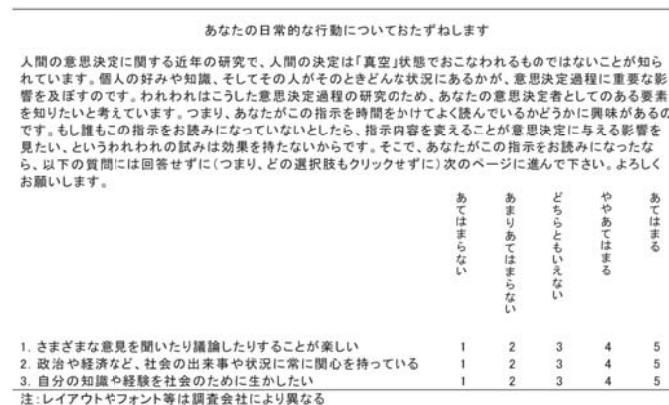


図2 三浦・小林(2015a)のIMC設問

まず、IMCでは対象としたA社では51.2%、B社では83.8%のサンプルが指示に従わないことが明らかになり、オンライン調査会社のパネル回答者の注意のレベルは概して低いことが明らかになった。また、IMC違反率はA社とB社の間で大きな差があり、いわ

ゆるHouse Effectがあることも示唆された。さらに、IMCで指示を遵守しなかった群（IMC違反群）は指示を遵守した群と比べて、後続のリッカート尺度における指示を無視する傾向が顕著に見られた（図3）。

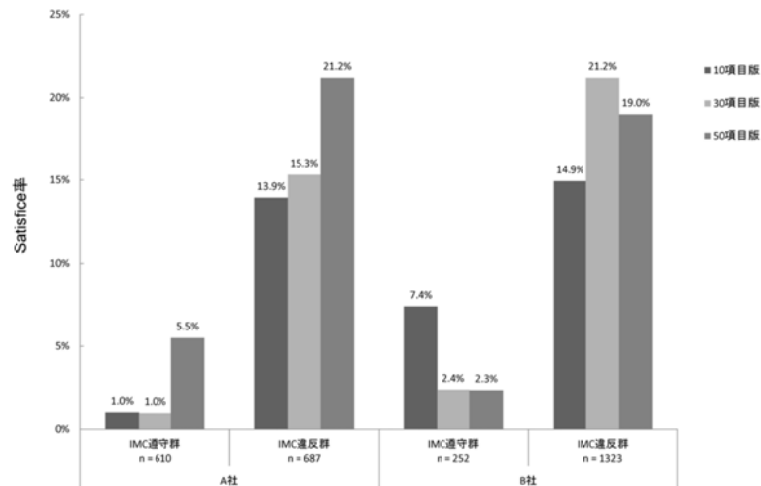


図3 リッカート尺度の読み飛ばし発生率（三浦・小林, 2015a）

三浦・小林(2015a)の結果は、日本のオンライン調査会社のパネルサンプルにおいて satisficeは広範に生じていること、またこうした回答者は回答の教示文や尺度項目を読まずに回答していることが強く示唆される。こうしたsatisficeはデータにおけるノイズ（測定誤差）を大きくするだけでなく、実験における処置効果の推定にバイアスをもたらすことが予想される。

三浦・小林(2016)は、三浦・小林(2015a)の結果をふまえて、satisficeが三浦・小林(2015a)で対象とした2社のオンライン調査会社を越えてどの程度一般的に見られるのかを検討した。その際、図2のIMCではなく、図4のものを用いた。

あなたの日常的な行動についておたずねします。

意思決定に関する近年の研究で、人間の決定は「真空」状態でおこなわれるわけではないことがわかってきました。人が何かを決めるとき、その人の好みや知識、または、そのときどんな状況におかれているかなどのさまざまな特徴が、大きな影響を及ぼすのです。この調査では、こうした「人間の決め方」を研究するために、あなたの「意思決定者」としてのある特徴を知りたいと考えています。つまり、あなたがこの指示を時間をかけてよく読んでいるかどうかを知りたいのです。もし誰もこの問題文をお読みになっていなければ、問題文の内容を変えることが「人間の決め方」に与える影響を見たい、というわれわれの試みは意味を持たないからです。ここからがお願いです。この指示をお読みになったことの証明として、実際のあなたがどうであろうが、以下の質問には「はい」と回答して、次のページに進んで下さい。よろしくお願します。

私は電子メールを使ったことがない

はい

いいえ

わからない

>>

図4 図2 三浦・小林(2016)のIMC設問

データ取得の対象集団は表1の通りである。三浦・小林(2015a)で対象とした2つのオンライン調査会社に加えて、オンラインクラウドソーシングサービスであるランサーズとクラウドワークス、および大学生サンプルとして関西学院大学の学部学生を対象とした。

表1 三浦・小林(2016)の調査協力依頼対象集団

調査	調査時期	調査協力依頼対象	IMC設問
1	2014年8月	ネット調査会社A	5件法3項目の無回答
2	2014年8月	ネット調査会社B	5件法3項目の無回答
3	2015年5月	ネット調査会社A	3肢1項目の選択肢指定
4	2015年12月	ネット調査会社A	3肢1項目の選択肢指定
5	2015年12月	クラウドソーシング(ランサーズ)	3肢1項目の選択肢指定
6	2016年1月	クラウドソーシング(クラウドワークス)	3肢1項目の選択肢指定
7	2016年1月	大学生参加者プール(関西学院大学)	3肢1項目の選択肢指定

各調査におけるIMC違反率は表2の通りであった。三浦・小林(2015a)の知見と同様、ネット調査会社Aでは40～50%の割合でIMC違反者が発生する。ネット調査会社Bにおける違反率は8割を越えている。一方、クラウドソーシングサービス2社と学生サンプルではIMC違反率は2割程度であり、オンライン調査会社のパネルサンプルよりも有意に

低いことが示された。これらのことから、日本におけるオンライン調査・実験はサンプルの注意のレベルという面から考える限り、クラウドソーシングサービスから参加者をリクルートすることが望ましいと考えられる。

表2 IMC 設問における教示違反率・遵守/違反者ごとの所要時間（三浦・小林, 2016）

調査	調査協力依頼対象	違反率(%)	所要時間（秒）		
			違反者	遵守者	差分
1	ネット調査会社 A	59.9	90.28(132.59)	134.45(158.88)	44.16**
2	ネット調査会社 B	82.5	91.78(136.16)	146.31(101.21)	54.53**
3	ネット調査会社 A	42.3	423.74(337.76)	495.73(303.90)	71.99**
4	ネット調査会社 A	44.2	251.02(221.55)	321.31(268.03)	70.29**
5	クラウドソーシング (ランサーズ)	19.5	388.70(277.65)	419.20(235.02)	30.50*
6	クラウドソーシング (クラウドワークス)	21.7	312.43(191.75)	378.19(200.42)	65.76**
7	大学生参加者プール	20.3	335.32(103.31)	373.61(139.58)	38.28+
カッコ内数値はSD					
** $p < 0.01$ , * $p < 0.05$ , + $p < 0.10$					

オンライン調査・実験の利点の1つは、データ収集の実施が距離や時間などの物理的制約から解放される点にある。たとえば、AMTを用いれば、日本国内にいつつも外国人を対象とした調査・実験が可能である。こうした利点を生かして、近年では国際比較実験がオンラインで実施されるようになりつつある。しかし、オンライン調査のパネル管理方法やクラウドソーシングサービスの質などは国によって大きく異なる可能性がある。こうした可能性を無視して国際比較データを分析すると、誤った結論を導く可能性がある。ここでは、報告者らが実施したオンライン国際比較実験の結果からIMCに関する部分を紹介し、注意のレベルは各国で利用するサービスによって大きく異なることを示す。

実験は2016年～18年にかけて実施され、各国における利用サービスとサンプルサイズは表3の通りである。各国において図4に示したIMCを実験の冒頭で実施した結果、違反率は図5の通りであった。

表3 国際比較実験の概要

	対象	N
中国	オンライン調査会社	800
ドイツ	クラウドソーシング	819
フランス	クラウドソーシング	803
香港	オンライン調査会社	821
インド	クラウドソーシング	830
日本	クラウドソーシング	812
韓国	オンライン調査会社	800
イギリス	クラウドソーシング	807
アメリカ	クラウドソーシング	822

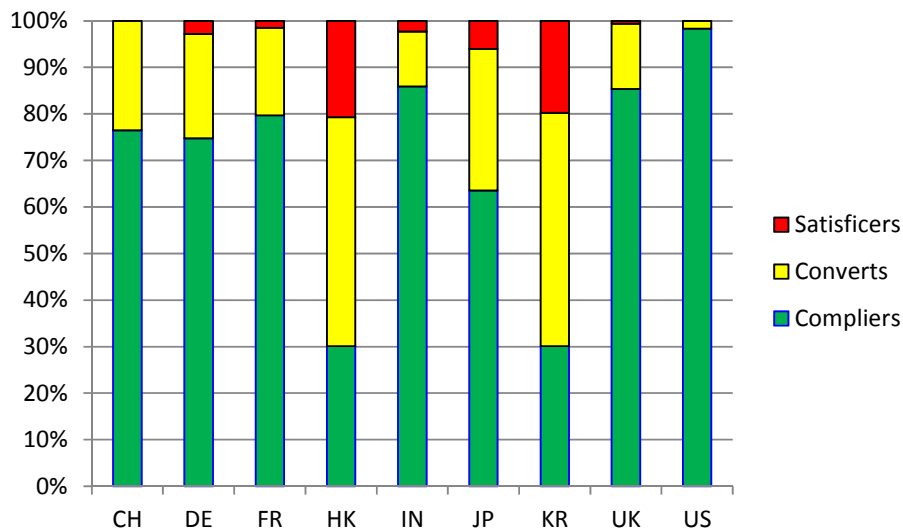


図5 IMC違反率の国際比較

各国において最初に提示されたIMCで違反が生じた場合には再度IMCが提示された。2回目のIMCでは回答の如何に関わらず次の設問に進むことが可能であった。これによって、回答者を3種類のタイプに分類することができる。まず、図5のcompliersは最初のIMCで指示を遵守した回答者であり、satisficeを行っていない「優良回答者」である。Convertsは最初のIMCで違反したものの、2回目のIMCでは指示を遵守した回答者である。彼らは当初は注意のレベルが低かったものの、IMCが繰り返されたことで注意のレベルを引き上げ、2回目で遵守することができたと考えられる。Satisficersは2回のIMC

の両方で違反した参加者であり、彼らは教示に対する注意のレベルが著しく低く、研究の知見を歪める危険性をもたらす。

図5から、国際的に見てもクラウドソーシングサンプルの方がIMCの遵守率が高いことがわかる。日本ではやや遵守率が低いものの、クラウドソーシングを用いた国では遵守率はすべて6割を越えている。一方、オンライン調査会社を用いた香港と韓国では遵守率は30%程度にとどまり、多くの参加者が教示に注意を払っていないことが明らかとなった。したがって、オンライン調査会社よりもクラウドソーシングサービス経由でリクルートされた回答者の方が概して注意のレベルが高いという傾向は日本国内だけでなく国際的にも一貫している。オンライン調査会社のパネルは各会社が管理しており、仮に信頼性の低い回答をしたとしても研究者自身がそうした行為を罰することはできない。しかし、クラウドソーシングの場合には研究者と各回答者が直接契約を結ぶため、もし信頼性の低い回答をした場合にはその参加者にネガティブな評判を付与することで罰することが可能である。つまり、satisficeが検出されるような信頼性の低い回答については報酬の支払いを拒否しネガティブな評判を付けることによって、そうした信頼性の低い回答者が今後ほかのタスクを請け負うことができにくくすることが可能である。回答者の側はネガティブな評判が付与されることによって将来獲得できる報酬が減るのを避けるので、satisficeをしない方向でのインセンティブがかかる。こうした評判システムが機能することで、クラウドソーシング経由の参加者の方が注意のレベルが高く、信頼性の高い回答をする傾向が生じると考えられる。

## Satisficeの弊害

Satisficeがオンライン調査会社のパネルサンプルを中心に広範に生じていることが明らかになったが、Satisficeは具体的にオンライン調査・実験の知見に具体的にどのような歪みをもたらすのだろうか。Satisficeが生じていても全体的な回答傾向がSatisficeを行わない回答者と大きく異ならないのであれば、Satisficeの弊害は大きくないだろう。しかし、以下に示すように、Satisficeを行う参加者の回答傾向は他の回答者と大きく異なっており、彼らを分析に含めることで仮説検証にバイアスが混入することとなる。



三浦・小林(2015b)は、オンライン調査会社（A社）のパネルを対象に、議題設定効果とプライミング効果を検証する実験を行い、Satisficeのレベルがこれらのよく知られたメディア効果の推定にどのように影響するのかを検討した。オンライン調査上で提示される動画の内容を操作し、処置群では「政治とカネの問題」に関連するニュース報道が提示される一方、統制群では提示されなかった。処置群で「政治とカネの問題」争点の重要度認知が高まれば、議題設定効果が確認されたことになる。また、処置後の安倍内閣の全体的な業績評価を従属変数とした場合に、処置前の「政治とカネの問題」に関する安倍内閣の業績評価と処置の間に正の交互作用が見られればプライミング効果が確認されたことになる。

ニュース動画の閲覧時間をバックグラウンドで計測し、動画の見飛ばしというsatisficeを測定した。優良群はニュース動画を最後まで閲覧した参加者であり、見飛ばし群は明らかに動画を最後まで見ずに次のページへ進んだ参加者である。最優良群は、優良群のうちニュースの内容に関する質問にすべて正答した回答者である。議題設定効果を検証した表4より、最優良群では有意ではないものの正の処置効果が推定されたが、見飛ばし群では仮説とは逆の負の効果が推定されており、見飛ばしのレベルによって処置効果の推定結果が異なることが示された。表5はプライミング効果の検証結果を示している。最優良群と優良群では仮説どおりプラスの交互作用が見られたが、見飛ばし群では仮説は逆の負の効果が推定されており、結果としてサンプル全体でプライミング効果を推定した場合には有意な正の交互作用が見られなくなっている。このことは、動画の見飛ばしを行うようなsatisficersから得られたデータが含まれている場合、本来検証されるべき効果が見られなくなることを示唆している。同様に、Miura & Kobayashi (2016)はオンライン調査会社のパネルを対象にヴィニエット形式のオンライン実験を実施し、IMCで同定されたsatisficersは他の参加者と比較してヴィニエットを読む時間が有意に短く、ヴィニエット内に描かれた仮想的な人物の印象評価を行う際により認知的コストの低いステレオタイプの判断をする傾向が強いことが示された。すなわち、satisficersは回答行動における認知的コストを最小化しようとするため、人物の印象評価においてもヴィニエットで提示された情報を逐次的に処理はせず、国籍などの周辺の手がかり情報に基づいて効率的なステレオタイプの判断を行う傾向があることが明らか

となった。こうした結果は、satisficeは仮説検証にバイアスをもたらす危険性があることを強く示唆している。

表4 議題設定効果の推定におけるsatisficeの影響（三浦・小林, 2015b）

従属変数: 「政治とカネの問題」争点の重要度認知 (処置後)	全体	見飛ばし群	優良群	最優良群
	Coef. (B)			
映像刺激の種類 (処置)	-0.178 (0.143)	-0.037 (0.182)	-0.089 (0.244)	0.109 (0.284)
報酬の種類 (抽選)	0.187 (0.143)	0.182 (0.303)	0.325 (0.245)	0.240 (0.288)
「政治とカネの問題」争点の重要度認知 (処置前)	12.571** (0.468)	10.024** (0.929)	13.866** (0.824)	14.283** (0.961)
ブロック変数 (1: 1波IMC遵守ブロック、2: 1波IMC違反ブロック)	-0.087 (0.144)	0.267 (-0.387)	-0.064 (0.250)	-0.083 (0.294)
カットポイント1	0.928** (0.249)	1.006 <sup>†</sup> (-0.611)	1.207** (0.430)	1.467* (0.677)
カットポイント2	4.179** (0.261)	2.952** (-0.619)	4.840** (0.468)	5.202** (0.718)
カットポイント3	7.671** (0.333)	5.966** (-0.747)	8.753** (0.615)	9.056** (0.857)
カットポイント4	10.952** (0.425)	8.808** (-0.892)	12.243** (0.781)	12.747** (1.034)
N	946	189	365	286
擬似決定係数	0.479	0.382	0.539	0.554

カッコ内数値は標準誤差  
 \*\* $p < 0.01$ , \* $p < 0.05$ , <sup>†</sup> $p < 0.10$   
 共変量は群ごとに異なるので省略した。

表5 プライミング効果の推定におけるsatisficeの影響（三浦・小林, 2015b）

従属変数: 現内閣の全体的評価 (処置後)	全体	見飛ばし群	優良群	最優良群
	Coef. (B)			
「政治とカネの問題」争点の評価 (処置前)	5.477** (0.383)	2.493** (-0.922)	4.971** (0.671)	4.405** (0.781)
映像刺激の種類 (処置)	0.151 (0.196)	0.759 <sup>†</sup> (-0.452)	-0.450 (0.316)	-0.515 (0.362)
「政治とカネの問題」争点の評価 (処置前) × 処置	-0.572 (0.484)	-1.666 (-1.095)	1.509 <sup>†</sup> (0.839)	1.962* (0.948)
報酬の種類 (抽選)	-0.094 (0.120)	0.142 (-0.28)	-0.382 <sup>†</sup> (0.196)	-0.267 (0.229)
ブロック変数 (1: 1波IMC遵守ブロック、2: 1波IMC違反ブロック)	0.196 (0.120)	0.565 (-0.364)	0.182 (0.197)	0.181 (0.229)
カットポイント1	-0.448** (0.165)	0.817* (-0.4)	-0.587* (0.275)	-0.166 (0.578)
カットポイント2	0.972** (0.166)	2.938** (-0.462)	0.681* (0.276)	1.025 <sup>†</sup> (0.576)
カットポイント3	2.614** (0.184)	5.460** (-0.567)	2.210** (0.300)	2.509** (0.591)
カットポイント4	5.052** (0.238)	8.215** (-0.719)	4.913** (0.394)	5.190** (0.674)
N	946	189	365	286
擬似決定係数	0.134	0.296	0.156	0.156

カッコ内数値は標準誤差  
 \*\* $p < 0.01$ , \* $p < 0.05$ , <sup>†</sup> $p < 0.10$   
 共変量は群ごとに異なるので省略した。

Satisficeの弊害は実験における処置効果推定のバイアスに限定されるものではない。オンライン調査を用いた心理尺度の測定においても、バイアスをもたらす。三浦・小林

(2018)は、オンライン調査会社とクラウドソーシングサービスから調査回答者をリクルートし、Directed Questions Scale (DQS; Maniaci & Rogge, 2014)によってsatisficeのレベルを測定した上でBig Five尺度や自尊感情尺度などのさまざまな心理尺度への回答を求めた。DQSは尺度項目を精読しない努力の最小限化を検出するための方法で、多数の項目からなるリッカート尺度に、選択すべき選択肢を指示する項目を含め、指示通りの選択がなされたかどうかによって判定する。本研究では、いくつかのパーソナリティ・態度を測定するリッカート形式の尺度に「\*\* この項目は「1. まったくあてはまらない」を選択して下さい」と回答する選択肢を指定する項目を含めた。その結果、DQS遵守群と違反群の間では因子構造に大きな違いが見られることが明らかとなった。たとえば、Big Five尺度因子への回答では、各因子を測定する2つの反転項目は負の相関を示すことが期待される。表6に示すように、調査AとBともに遵守群では予測どおり各因子における2項目は負相関を示しているのに対して、違反群では多くの場合正相関を示している。これはDQSで検出されるsatisficersは尺度の項目を読み飛ばしていることを強く示唆している。

表6 Big Five尺度因子の項目間相関（三浦・小林, 2018）

因子	項目	小塩ら (2012)	調査 A		調査 B	
			遵守 (n=5318)	違反 (n=842)	遵守 (n=1173)	違反 (n=58)
外向性	活発で、外向的だと思う	-0.59	-0.44	0.08	-0.53	-0.05
	ひかえめで、おとなしいと思う					
協調性	他人に不満をもち、もめごとを起こしやすいと思う	-0.22	-0.20	0.17	-0.19	0.16
	人に気をつかう、やさしい人間だと思う					
勤勉性	しっかりしていて、自分に厳しいと思う	-0.38	-0.37	0.20	-0.42	-0.14
	だらしなく、うっかりしていると思う					
神経症傾向	心配性で、うろたえやすいと思う	-0.28	-0.30	0.18	-0.31	0.15
	冷静で、気分が安定している					
開放性	新しいことが好きで、変わった考えをもつと思う	-0.39	-0.28	0.17	-0.37	0.19
	発想力に欠けた、平凡な人間だと思う					

また、三浦・小林(2018)は自尊感情尺度の検証的因子分析から、DQS遵守群は反転尺度の因子負荷の符号は予測どおり反転するのに対して、違反群は反転項目も含めてすべての項目において正の因子負荷が推定されるなど、解釈不能な因子構造を示すことを報

告している。さらに、二分法的志向尺度の探索的因子分析では、DQS遵守群では2因子構造が見られる一方で、違反群では1因子構造が見られるといった違いも報告されている。こうした結果は、satisficersは尺度項目を十分に注意して読んでおらず、そのことが解釈不能な項目間相関や因子構造をもたらしていると解釈できるだろう。こうした信頼性の低い回答を含めて分析した場合に、本来検出されるべき因子構造が得られなかったり、分析に用いたとしても測定誤差が大きいために相関の希薄化が生じて仮説が適切に検証されないといった弊害が考えられる。このように、オンラインにおけるデータ収集で発生するsatisficeを含めて分析を行うと、実験の処置効果の推定にバイアスがかかったり、調査における尺度の測定に失敗するなどの望ましくない結果を招く危険性がある。では、こうしたオンライン調査・実験で発生するsatisficeに対してどのように対処するのが望ましいのだろうか。

## Satisficeへの対処

オンライン調査・実験では研究者が回答行動を直接監督することができないため、間接的な方法で参加者の注意のレベルを測定する必要がある。そのために有効な手段は大きく分けて2つある。1つは本稿で紹介した研究で繰り返し用いられているように、Instrumental Manipulation Check (Oppenheimer et al., 2009)やDirected Questions Scale (DQS; Maniaci & Rogge, 2014)などを用いて、調査票の中に注意のレベルを測定する項目を盛り込む方法である。これらの方法は教示や刺激を読み飛ばしていると「求められた回答」ができないことを利用して注意のレベルを測定することを可能にする。2つ目は、回答者が各ページに滞在した時間などのメタデータをバックグラウンドで測定することである。クラウドソーシングを用いて参加者を募集する場合にはQualtricsを用いて研究者自身が調査・実験用のサイトを構築するのが一般的である。この際に、回答者が用いているデバイスの種類や回答にかかった時間などをバックグラウンドで測定することが可能である。こうした測定は、IMCやDQSのように明示的に調査に含めて回答者の注意を測定する方法とは異なり、非侵襲的に測定できることが特徴である。たとえば、回答時間

が極めて短い参加者を事後的に分析から除外するなどの対処が可能になる。近年ではオンライン調査会社でも研究者がQualtricsで構築したサイトに自社のパネル登録者を誘導することを許可する会社が増えてきているため、必ずしもオンライン調査会社に依頼せずとも自前でこうしたメタデータを取得することのハードルは低くなりつつある。

IMCやDQSを用いた注意レベルの測定やメタデータの取得は、いわば事後的にsatisficeに対処することを可能にする。しかし、そもそもsatisficeの発生そのものを抑制することはできないのだろうか。これについては、可能であるという立場の研究(e.g. Oppenheimer et al., 2009)と、satisficerに注意のレベルを高めるように促してもデータの質は変わらないとする立場の研究(e.g. Berinsky et al., 2016)がある。Satisficeの発生にはこれまでに見てきたように国や参加者をリクルートするソースの影響があるため、日本での知見を得る必要がある。ここでは、Miura & Kobayashi (2016)の研究を紹介し、IMCを繰り返すことでsatisficerの注意のレベルは部分的に高められることを示す。

Miura & Kobayashi (2016)は、既に述べたように、オンライン調査会社のパネルを対象にヴィニエット形式のオンライン実験を行った。その際、IMCを実験の冒頭で最大2回繰り返して測定した。最初のIMCで指示を遵守した回答者(compliers)は2回目のIMCは提示されなかった。一方、最初のIMCで違反した者は、2回目のIMCで指示を遵守した回答者(converts)と2回目のIMCでも違反した回答者(satisficers)に分類された。ここで注目するのはconvertsである。Convertsの多くは、1回目のIMCでは注意のレベルが低く指示に違反したが、次の設問で同じIMCが再度提示されたことで注意のレベルが高まり、指示を遵守することができた回答者である。問題は、後続の実験においてconvertsがcompliersと同じような振る舞いを見せるかどうかである。もし、繰り返しIMCによって高められた注意のレベルが後続の実験でも維持されるのであれば、convertsはcompliersと同じような振る舞いを見せるだろう。一方、IMCによって一時的に高められた注意のレベルがすぐに低下してしまうのであれば、convertsはむしろsatisficersに近い振る舞いを見せるだろう。図Xは3種類の回答者ごとに実験の効果を示したものである。横軸はヴィニエットとして提示された仮想的な人物に対する印象評定（右の方が好印象）であり、人物の国籍（日本または中国）および人物の特性（暖かい・冷たい・記述なし）の6条件ごとの結果が示されている。

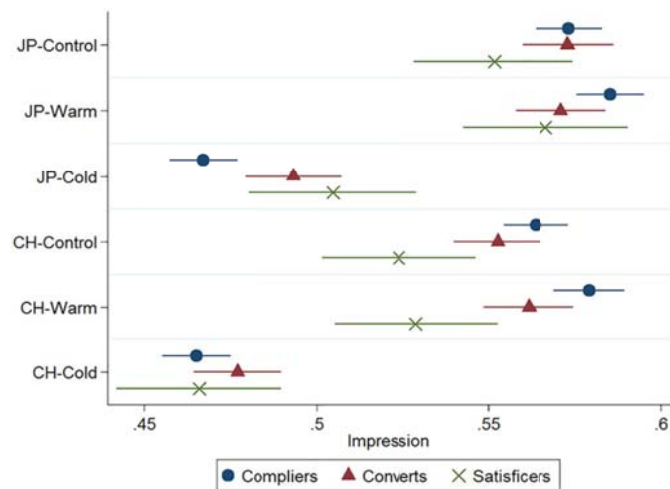


図6 繰り返しIMCによる注意レベルの引き上げの効果 (Miura & Kobayashi, 2016)

図6から明らかなように、convertsの結果はsatisficersよりもcompliersに近づく場合が多く、Convertsとcompliersの間に有意な差が見られたのはJP-Cold条件のみである。ただし、convertsの結果は完全にcompliersのそれと一致するわけではなかった。傾向としてはcompliersに近づくが、compliersよりはややsatisficersに近い傾向を示す。この傾向は、Miura & Kobayashi (2016)の実験2でも一貫していた。すなわち、IMCを繰り返すことで注意の低さを自覚した潜在的なsatisficersは注意のレベルを引き上げ、完全ではないがcompliersに近いレベルで後続の実験刺激にも注意を払うことが示された。このことは、IMCはsatisfice傾向を測定するだけでなく、繰り返すことで潜在的なsatisficeを抑制することができることを示している。

## 結論

オンライン調査・実験は簡便かつ低廉に多様なサンプルを対象に調査や実験を行うことが可能なため、今後も広く使われ続けるだろう。しかし、回答行動を直接観察できなかつたり、回答者のインセンティブ構造によってsatisficeが広範に発生する危険性が付きまとう。本稿で紹介した一連の日本での検証は、日本でもオンライン調査会社のパネルを中心にsatisficeは広範に生じていること、satisficeは処置効果の推定を歪めたり尺度

の測定に無視できないノイズを混入させる危険性があること、しかし適切な測定項目やメタデータを用いることでsatisficeの発生を把握し、部分的には抑制できることを明らかにしている。今後はオンラインでデータを収集する際には、何らかの方法でsatisficeに対処することが強く求められるだろう。

## 引用文献

- Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2016). Can we turn shirkers into workers?. *Journal of Experimental Social Psychology*, 66, 20-28.
- Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, 48, 61-83.
- Miura, A., & Kobayashi, T. (2016). Survey satisficing inflates stereotypical responses in online experiment: The case of immigration study. *Frontiers in Psychology*, 7:1563.
- 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(1), 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(2), 120-127.
- 三浦麻子・小林哲郎 (2016). オンライン調査における努力の最小限化 (Satisfice) 傾向の比較: IMC 違反率を指標として メディア・情報・コミュニケーション研究, 1, 27-42.
- 三浦麻子・小林哲郎 (2018). オンライン調査における努力の最小限化が回答行動に及ぼす影響. 行動計量学, 45(1), 1-11.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4), 867-872.