

行動研究における新しい潮流

—クラウドソーシングとオンライン・サーヴェイプラットフォームの連携—

眞 嶋 良 全

行動研究における新しい潮流

——クラウドソーシングとオンライン・サーヴェイプラットフォームの連携——

眞 嶋 良 全

Yoshimasa MAJIMA

目次

1. 行動研究におけるオンラインの実験・調査環境
 - 1.1 クラウドソーシング
 2. オンライン研究に伴う諸問題
 - 2.1 クラウドソーシング導入のメリット, デメリット
 - 2.2 データの信頼性, 妥当性
 - 2.3 技術的問題
 - 2.4 倫理的問題
 3. 日本におけるクラウドソーシング・サンプルを用いたオンライン調査・実験環境の構築へ向けて
 - 3.1 クラウドソーシング・サービスとオンラインのサーヴェイプラットフォームの併用
 - 3.2 日本における導入にあたって考慮すべき問題
- 引用文献

[Abstract]

Recent Trends in Behavioral Research Methods: Cooperative Use of Crowdsourcing Services and Web-based Survey Platforms

Online crowdsourcing services are a relatively new labor portal where anonymous workers earn small amounts of money for completing Web-based tasks. Recently, social scientists are paying attention to this attractive alternative to traditional university student participant pools for the purpose of collecting survey and experimental data in behavioral research. One reason is that using a crowdsourcing service together with an online survey platform helps researchers conduct behavioral surveys and experiments quickly and efficiently. The present article reviews recent studies examining the quality, validity and reliability of data from online workers. These previous studies indicate that data from crowdsourcing samples are mostly reliable; however, some demographic characteristics are different from traditional student samples. The present article also discusses some technical and ethical issues in using crowdsourcing services as a participant pool and Web-based survey platforms in behavioral research in Japan.

1. 行動研究におけるオンラインの実験・調査環境

心理学, および隣接領域では, その理論・仮説の検証のために, 人の行動をデータとする行動研究 (behavioral research) が行われる。そこで取得される行動は, 極めて多様であり, 広義の行動は, 外部から観測可能な, 生体からのあらゆる出力を含んでいる。例えば, 生理心理学で用いられる自律神経系に支配された生理活動や, 神経心理学の領域で用いられる脳活動も広い意味では行動の一部と考えることができる。さらに, 体性神経系に支配される骨格・筋肉系の運動反応や, 言語

的に表現される諸反応, 例えば, 思考内容の発話, 記憶における再生, 人格尺度や態度等についての自己報告等も行動の一種である。行動研究においては, これらの多様な行動を測定するために, さまざまな調査や実験を行う。その中でも, 特に紙と鉛筆を用いる質問票 (paper-and-pencil task) による反応の取得は, その実施の簡便さもあって, 多くの行動研究で用いられている。

通常, この種の質問票は, 研究者が所属する大学での講義時間等を利用して配布されることが多く, 参加者はその場で質問票に回答し, その結果を表計算, あるいは統計ソフトウェアに入力した後で, 分析が行われる。大

キーワード: クラウドソーシング, 行動研究, ウェブベースの研究デザイン

Key words: Crowdsourcing, Behavioral Research, Web-based Research Design

規模な調査の場合は、質問票を郵送するなどして参加者募集にまつわる労力を軽減したり、マークシートを利用して入力の手間を省くことによって時間・労力面でのコストを削減することもある（ただし、その場合は金銭面でのコストは増大する）が、いずれの場合も調査を計画してから、データを取得し分析が完了するまでに多くの時間と費用が必要になる。参加者を個別に呼んで実験室で行う認知実験の場合は、その完了に必要な日数は調査よりもさらに長くなる。

このような研究の効率を低下させる諸要因、特に労力・時間面でのコストを軽減するための一つの解決策が、近年爆発的に普及したPCやスマートフォン等の情報端末を利用した、オンラインのWebベースの研究環境（以下、サーヴェイプラットフォームと呼ぶ）の導入である。オンラインでの調査・実験が行われるようになったきっかけは、Web技術の革新により、サーバとユーザの相互作用が可能になったことにある。後述するクラウドソーシング・サービスが導入される以前から、オンライン上で参加者を募集して行う調査や実験の試みがなされており、例えば、1998年の時点のAPS（American Psychological Society）によるリストでは、オンライン上の心理学研究プロジェクトは35件が登録されていたが、1999年の時点でその件数が65まで増えている（Birnbbaum, 2000）¹。しかし、この時点では、データを実際に取得するサーヴェイプラットフォームをオンライン上に構築する際の技術的側面が問題となることが多く、Webサイト上で募集したり、E-Mail等を利用する試みは行われていたものの、参加者の募集を効率的に行う仕組みはまだ整備されていなかった。しかし、近年では、クラウドソーシング（crowdsourcing）という、オンライン上で業務を外部委託するシステムの整備に伴い、そのサービスを行動研究においても使おうとする試みがなされるようになってきて

いる。

1.1 クラウドソーシング

クラウドソーシングは「不特定の人（クラウド、群衆）に業務を外部委託（アウトソーシング）するという意味の造語であり、発注者がインターネット上のウェブサイトで受注者を公募し、仕事を発注することができる働き方の仕組み（総務省, 2014, p.210）」であるとされる。欧米を中心に普及が進み、日本国内でも、その市場規模は2013年度では250億円弱、2017年には1000億を超えると推定されている（総務省, 2014）。多くのクラウドソーシング・サービスでは、受注可能な仕事を、プロジェクト、コンペ、タスクという大まかなカテゴリに分けている。このうち、プロジェクトは相対的に大型の案件について、事前に設定された要件に基づいて一連の課題を達成し、その成果に対して一定額の報酬が支払われるもの、コンペは設定された課題に対して複数の受注者が作成した成果物の中から、発注者が一つを選び、選ばれた受注者に対して報酬が支払われるものを指すことが多い。一方タスクは、短いものでは数分、長くても数時間程度で完了するような小規模な課題（例えば、発注者が依頼する単純な調査や、単純なデータ入力、短い記事やコンテンツのライティングなど多岐にわたる）が多く、報酬額も1件あたり数円から数千円程度と少額なものが多い。

多くのクラウドソーシング・サービスでは、プロジェクト、コンペを中心に据えているものが多いが、タスクに特化したクラウドソーシング・サービスもある。そのようなタスク特化型のクラウドソーシング・サービスの代表格とも言えるのが、米国Amazon社が展開するサービスのAmazon Mechanical Turkである（以下、AMTと書く）。Mechanical Turk（機械仕掛けのトルコ人）という名称は、18世紀にハンガリーのWolfgang von

Kempelen が作成したチェス対戦ロボットからきている。このロボットはほぼ全ての挑戦者に勝利することができたが、実は中に本物のチェスの名人が隠れていた。この故事にちなみ、AMT は、コンピュータが対処できない大量の、しかし小さな問題を人力で解決することを目的に開発された (Terdiman, 2005)。

AMT において、作業の受注者は Worker と呼ばれ、多数の HITs (Human Intelligence Tasks) というタスクの中から希望する HITs を請け負う。作業を依頼する発注者は Requester と呼ばれ、Worker が依頼を完遂した時に報酬を支払う。Requester は受注者全員の応募を承諾し、成功報酬を支払う必要はなく、基準に満たない Worker からの応募を事前に拒否したり、達成度に問題のある Worker に対しては報酬を支払わないという選択も可能である。本来は、コンピュータには困難だが人にとっては容易な小規模のタスクについての労働力をオンライン上でやり取りするために開発された労働契約ポータルであるが、近年では、その多数かつ多様な Worker²を、大規模な調査・実験を高速かつ効率的に行うための参加者プールとして利用するケースが、心理学をはじめとする社会科学において増えてきている (Behrend, Sharek, Meade, & Wiebe, 2011; Buhrmester, Kwang, & Gosling, 2011; Chandler, Mueller, & Paolacci, 2014; Mason & Suri, 2012)。

AMT をはじめとするクラウドソーシング・サービスを行動研究において利用する利点は、単に多数の参加者を効率的に集めることができるということに留まらない。クラウドソーシング・サービスは、発注者と受注者がオンライン上で労働力と、それに対する対価 (報酬) をやり取りするサービスであるため、タスクの発注→受注→タスクの達成→報酬の支払い、という一連の流れ、すなわち行動研究においては、参加者の募集から調査、実験の

実行、協力に対する謝礼の支払いまでが、一つのクラウドソーシング・システム内で完結する³。このように、研究を遂行するにあたって必要な諸手続きのかなりの部分が、一つのシステム内で行えることも、労力面でのコスト軽減に一役買っている。

行動研究のための調査や実験の HITs は、その所要時間にもよるが、1件あたりの報酬はせいぜいで数百円程度である。例えば、1時間程度の HITs の場合は1.4ドル、5分程度と短い HITs の場合は5セントほどの報酬が支払われるケースが多い⁴。AMT で行動研究タスクを請け負った Worker は、多くの場合、AMT の外部にあるサーヴェイプラットフォームへと誘導される。AMT にも簡単なアンケート程度であれば処理することのできるインターフェースが用意されているが、行動研究のツールとしての性能が十分とは言えず、実際の調査・実験は、その目的に特化した専門の Web サービスが用いられる。例えば、SurveyMonkey (<http://www.surveymonkey.com>), Qualtrics (<http://www.qualtrics.com>) 等は、よく行動研究に用いられる有償のサーヴェイプラットフォームである。また、Web サーバは別途用意する必要があるが、オープンソースの Web ベースのサーヴェイアプリケーションである LimeSurvey を利用するという方法もある (<http://www.limesurvey.org>)。

2. オンライン研究に伴う諸問題

2.1 クラウドソーシング導入のメリット、デメリット

行動研究においてオンライン上にサーヴェイプラットフォームをおくことのメリットは、極言すると大量のデータを短期間にかつ効率的に収集することを可能にするという点に尽きる。必要なサンプル数の大小によるものの、実験や調査を開始してから、早ければ数日で

数百件にわたるデータを収集することも可能である。このように、データ収集に必要な時間が短く、また実験・調査自体がオンライン上で完結するために、その実行中であっても研究者が立ち会う必要がないということ、さらに、調査・実験をサーバ上に実装するためのプログラムコードを研究者間で共有できることから、従来型の調査・実験の際に研究者が割かなければならなかった時間を、理論的な考察やそれに基づいた新たな研究デザインの質の向上に充てることができるようになる。

クラウドソーシング・サービスを経由して参加者を集めることは、多くの心理・行動研究にありがちな大学生サンプルに比べてサンプルの多様性が高まり、研究結果の妥当性が高まることに繋がるという利点もある。例えば、Behrend et al. (2011) によると、AMT 経由で参加したサンプルは、性別・人種の構成比については大学生サンプルと差はないが、教育、雇用状況、専門の分散が多様であり、平均年齢が大学生よりも高く、多くの参加者が被雇用者であり、在職年数も長かった。このことから、特に組織心理学的研究においては、クラウドソーシング・サービスの利用を推進するべきであることが指摘されている。日本において行動研究にクラウドソーシング・サービスを導入しようとする試みは、これまでのところ、あまり活発であるとは言えないが、総務省 (2014) によると、クラウドソーシング・サービスの登録者の性質として、30代の女性が多いものの、50歳以上のシニア世代が増加しつつあること、登録者数については、12,000人を越える規模に達していることなどが報告されている。行動研究では、しばしば、実験・調査の参加者を大学生に限定していることが結論の一般化可能性を妨げる要因であるという批判がなされることがあり、より妥当性の高い知見を得るためにも、多様なサンプルからのデータ取得を可能とするクラウドソーシング・サービスの利用を進める

ことが重要になってくると思われる。

さらに、行動研究においてしばしば問題となるアーチファクトである、実験実施者と対面で実験・調査を行う際に生じるさまざまな「実験者効果」を最小限に抑えることができる点も利点として挙げられている (Crump, McDonnell, & Gureckis, 2013)。

しかしながら、その一方で、オンラインの実験・調査においてクラウドソーシング・サンプルを用いることにはいくつか考慮しなくてはならない点や、場合によってはデメリットがあることも指摘されている。以下で、これまでに指摘されているいくつかの注意を要する点について述べる。

2.2 データの信頼性, 妥当性

AMT は既に多くの研究で、実際に参加者募集の手段として用いられている。しかし、オンライン上の匿名の参加者から得られたデータが、果たして母集団の性質を反映した代表的サンプルといえるかどうかには注意しなくてはならない。

行動研究において、参加者のリクルート手段として AMT を用いた場合のデータの信頼性については、性格や認知スタイル (Behrend et al., 2011; Buhrmester et al., 2011; Goodman, Cryder, & Cheema, 2013), 経済ゲーム (Amir, Rand, & Gal, 2012), ヒューリスティクス判断 (Goodman et al., 2013; Paolacci, Chandler, & Ipeirotis, 2010), 社会的ジレンマ (Suri & Watts, 2011), 言語 (Schnoebelen & Kuperman, 2010; Sprouse, 2011), 注意・知覚的判断 (Crump et al., 2013) 等、さまざまな調査・実験課題を用いた検討がここ数年の間行われている。

例えば、Behrend et al. (2011) は、人格尺度であるビッグファイブを用い、AMT 経由で募集したサンプルと大学生サンプルを比較し、両者は外向性以外の4尺度の得点に差

がなく、外向性についても検定結果は有意であったものの効果量が低いという結果を得ている。

また、AMT サンプルを報酬額（3群）、想定される所要時間（3群）の組み合わせによる9群に分け、そのそれぞれで人格尺度への回答を求めたところ、報酬が低い、または想定所要時間が長いほど調査自体への参加率は低下するものの、それらの群間で回答に差はなく、3週間後に行われた再検査に際しても十分に高い再検査信頼性（平均して $r = .88$ ）を保つことも示されている（Buhrmester et al., 2011）。

さらに、ストループ干渉、タスクスイッチング、フランカー課題、サイモン効果、注意の復帰抑制など、知覚や注意研究における古典的な実験を AMT サンプルを用いてオンラインで実施した場合も、先行研究通りの結果が再現されることが示されている（Crump et al., 2013）。これらの結果は、クラウドソーシング・サービス経由で収集したデータは十分に信頼に足ることを示している。

2.2.1 重複参加

しかしながら、一方で、AMT サンプルのデータの信頼性を脅かす可能性のある要因もいくつか指摘されている。その一つは、重複参加の可能性である。AMT では、各調査・実験がタスクとして公開されるが、同一の Worker が同じタスクを複数回受注することも可能である。有償のサーヴェイプラットフォームを使用したり、自前で用意した調査ツールの中でチェック機能を実装することによって、重複参加をある程度防ぐことは可能であるが⁵、研究者が過去に発注した別の課題、あるいは異なる研究者が発注した類似の課題を受注した経験のある Worker の参加を防ぐことは簡単ではない。このようなタイプではない参加者からのデータが含まれている場合、データに歪みが発生することが考え

られる。参加者のタイプ性が問題となる研究の場合は、類似の研究に既に参加したことがある Worker を除外する必要がある。特に、異なる研究者の類似の課題に参加したかどうかのチェックについては、研究者間での情報共有も必要となってくる（Chandler et al., 2014）。

2.2.2 Worker 間の情報共有

2つ目の要因は、Worker 間の情報共有の問題である。AMT の Worker は、実際にオンライン上で情報交換を行っており、特に、HITs に対する報酬の額や、未到達として報酬が支払われなくなる基準が何かについての議論が行われているが、一方で、そのやりとりは最小限に抑えられており、課題の内容が事前にリークされるケースは少ないことも指摘されている（Chandler et al., 2014）。上記の点からは、情報交換自体によるタイプ性の低下は必ずしも生じないかもしれないが、タスクレベルに比べて高額すぎる、あるいは少額すぎる報酬を設定した場合、その情報が広まることで Worker のタスク受諾や、タスクからの脱落に影響し、結果としてデータの代表性が崩れ、歪みが生じる可能性が指摘されている。

2.2.3 脱落率

3つ目の要因は、匿名の参加であるがゆえの脱落率の高さの問題である。一般に、Webベースの調査は、大学生を用いた伝統的な実験室実験に比べて中途での脱落率が高いことが指摘されている（Birnbaum, 2004; O'Neil & Penrod, 2001）。また、脱落率は、金銭報酬の多寡によっては影響されないが、報酬をどのように支払うかには影響され、例えば報酬の受け取りに個人情報が入力が必要になると脱落率が高くなることが指摘されている（O'Neil & Penrod, 2001）⁶。さらに、参加者が調査に参加する曜日としては、週末に参

加する方が、脱落率が低くなることも示されている。脱落率を抑えるための方法としては、実験的操作による群の割り当てをすぐには行わず、群分けをせずとも可能な簡単な質問をウォームアップとして行う方法や、逆に開始以前の段階で動機づけを下げる高いハードルを設定する方法 (Birnbaum, 2004)、実験・調査自体を参加者にとって魅力が感じられるようなものにする重要性も指摘されている (Crump et al., 2013)。しかし、そのような工夫をこらしたとしても、大学生サンプルに比べて脱落率が高くなることは避けられないと考えられる。そのため、可能な限り脱落率を抑える工夫をするだけでなく、成果の公表に際して具体的な脱落率を明記することが推奨されている (Crump et al., 2013; Reips, 2002 等を参照)。

この他にも、匿名であるために人口統計変数を取得しにくいことや、伝統的な実験室実験に比べて、実験環境の統制が困難であること、さらには、参加者は人間であることが想定されているが、ごく稀に人間ではない Worker (例えば、入力を自動化する bot) のデータが紛れ込んでいる可能性が否定できないことも問題点として挙げることができる。しかし、これらの点については、可能性はあるものの実用上はそれほど大きな問題とはならない、あるいは技術的に回避可能であることも指摘されている (e.g. Crump et al., 2013)。

2.3 技術的問題

2つ目に考慮すべき問題は、Web ベースの実験・調査を行うことに伴うさまざまな技術的困難である。ここでは、大きく分けて3つの観点から述べる。

2.3.1 正確な時間制御

まず第一に挙げられるのは、行動研究の中でも、特に知覚、注意、記憶の実験で行われ

る反応の取得と刺激の呈示に関連した問題、すなわち正確な時間制御である。紙による質問票を Web ベースの調査に置き換えた場合はこの種の制御は必要ないが、知覚、注意、記憶等の領域の実験室実験では、しばしば刺激が1秒未満というごく短い時間で呈示され、さらには、刺激が呈示されてから反応が行われるまでの反応時間もミリ秒単位で測定される。残念ながら、サーバと参加者の用いるクライアント端末との間の回線速度やネットワークの混雑状態がボトルネックとなり、HTMLのみを用いた場合は、このようなミリ秒単位の制御が難しい。しかし、ミリ秒単位での時間制御が可能な Javascript など、クライアント側で動作するプログラム言語を用いることで、そのような問題を解決することができる (Birnbaum, 2004)。Crump et al. (2013) は、ストループ干渉、タスクスイッチング、フランカー課題、サイモン課題、Posner の先行手がかり課題、注意の瞬き、視覚的プライミング課題など、知覚や注意の領域における古典的で頑健な現象についての実験課題を、Javascript を用いた Web ベースの課題として実装し、AMT 経由で募集した参加者から取得した反応データが先行研究と同様の傾向を見せるかどうかを検討した。その結果、AMT サンプルの結果は、先行研究をほぼ再現していることが明らかになった。従って、Web ベースの実験環境においても、ミリ秒単位での反応の取得については概ね問題なく可能であると考えられる。

しかし、閾下プライミングなど、さらに短い時間での刺激呈示を要する実験を Web ベースのサーヴェイプラットフォームにおいて実行可能かどうかについての検討などは今後も進める必要があると思われる。加えて、このような認知実験を実装するためには、Javascript を始めとするプログラミング言語に精通する必要があるため、不慣れな研究者にとっては少々ハードルが高いかもしれない。実験

室実験でこのような実験課題をプログラミングする場合は、有償の実験環境構築ソフトウェア (SuperLab, E-Prime 等が有名である) を用いることで、プログラミングの経験がなくても比較的簡単に実験課題を作成することができるが、例えば、そのようなソフトウェアのうち、Millisecond Software 社の Inquisit (<http://www.millisecond.com>) は、PC にインストールして使用するクライアント版とは別に、Web サーバ上で実験を実装できるサービスを展開している (ただし、ライセンス料が年額で \$1,500 弱であり決して安価とはいえない)。また、Javascript を用いた行動実験のための共有ライブラリとして jsPsych (de Leeuw, 2014) というプロジェクトがあり、そこで公開されているライブラリを利用すれば、実験環境導入の手間を多少は軽減することが可能であろう⁷。

2.3.2 クラウドソーシング・サービスとサーバプラットフォームの連携

二点目の技術的問題は、クラウドソーシング・サービスとは別に、外部のサーバプラットフォームを用意する際に、両者の連携をどう取るのかという問題である。例えば、多くのクラウドソーシング・サービスでは、登録者には一意の ID が割り当てられていることがほとんどであり、その ID を取得することができれば、同じ調査・実験への重複参加を防ぐことができる。AMT においても、そのような機能を用いることは可能であるが、AMT では、高度な機能を使用するためには、Web ブラウザからの操作ではなく、コマンドラインからのコマンド入力による制御が必要になる (AMT の利用に伴う技術的諸問題は、Mason & Suri, 2012 が詳しい)。前述の実験プログラミングと同様に、この点も不慣れた研究者にとってはハードルが高くなる一因である。しかし、AMT を用いた研究を

サポートするツールを用いることで多少はハードルを下げることもできるかもしれない⁸。

2.3.3 データ・セキュリティ

三点目は、後述する倫理的問題とも関連するが、オンライン上で取得したデータの保護に代表されるセキュリティ上の問題である。オンライン上で行われる調査・実験では、個人を特定できる情報 (氏名、住所等の連絡先) は取得すべきではないが、それらの個人特定情報を得なかったとしても、行動研究では人格や、知的能力等のセンシティブ情報を扱うことが多い。そのため、データの保護には特に注意を払う必要がある。クラウドソーシング・サービス、あるいは有償のサーバプラットフォームにおいては、通常、個人情報保護のポリシーが設定されており、またそれを実現するためのセキュリティが導入されているのが普通であるが、サーバプラットフォームを研究者が用意する場合においては、そのサーバ環境でのデータの保護に万全を期する必要がある。特に、Web サーバ上のデータベースに対する SQL インジェクション攻撃や、クロスサイトスクリプティング (XSS) 攻撃等によるデータの漏洩に注意し、加えて通信経路を暗号化するなどの措置を講ずる必要がある。もちろん、サーバプラットフォームからダウンロードしたデータの保護にも十分な配慮が必要である。

2.4 倫理的問題

オンラインでの調査・実験を行うにあたって考慮すべき 3 つ目のポイントは、研究倫理に関連した諸問題である。

2.4.1 個人情報と研究データの保護

まず、前節でも述べたように、調査・実験の過程で得た全てのデータや、情報は適切に保護されなければならない。しかし、この個人情報の収集・保護、および研究データの管

理は、Webサーバ上にデータが置かれることに起因する固有の技術的問題を除けば、従来型の大学生サンプルを用いた研究についても当てはまる一般的な研究倫理でもある。一方で、AMTのようなクラウドソーシング・サービスと外部のサーヴェイプラットフォームの併用は、データ収集のプロセスと参加者管理の機能を分離することにつながる。参加者管理を担うクラウドソーシング・サービスでは、多くの場合、個々の参加者は個を識別できるが、人物特定情報とは切り離されたIDによって管理され、タスクを発注する研究者の側からは個人を特定できる情報を入手する必要がない（というよりは入手できない）。そのため、通常の実験室実験等と比べて、参加者の個人情報の扱いについてのハードルがやや低くなるという側面もある。しかしながら、タスク完了後のフォローアップや、参加者の側から研究成果についての問い合わせがあった場合、連絡先としてのE-Mailアドレス等を入手する機会もあると思われる。その情報の保護については、当然のことながら配慮する必要がある。

2.4.2 インフォームド・コンセント

研究倫理について、もう一つ考慮を要する重要な問題は、インフォームド・コンセントである。研究実施前のインフォームド・コンセントについては、日本心理学会倫理規程では以下のように定められている（日本心理学会、2009）。

“研究にたずさわる者は、研究対象者に対し、研究過程全般および研究成果の公表方法、研究終了後の対応について研究を開始する前に十分な説明を行い、理解されたかどうかを確認した上で、原則として、文書で同意を得なければならない。説明を行う際には、研究に関して誤解が生じないように努め、研究対象者が自由意志で研究参

加を決定できるよう配慮する。(p.11)”

オンラインでの調査・実験の場合、このインフォームド・コンセントはWeb上のテキストで行うことになるが、研究者との接触が制限されているため、疑問点をその場で解消することができない点には注意する必要がある。なお、Webサーバ上で行われるインフォームド・コンセントに際して、説明文書が必ずしもきちんと読まれていないという指摘もある（Behrend et al., 2011; Stanton & Rogelberg, 2001）。

2.4.3 虚偽の説明とディブリーフィング

心理学の中でも、特に実験的研究では、事前に研究の真の目的を伝えることが、参加者の反応を変化させてしまうことがある。そのようなケースでは、事前に倫理委員会の承認を得た上で、虚偽の説明による実験を実施することが容認されている。この場合も、遅くとも研究終了時点で参加者に対して虚偽の説明があったことを伝え、研究の真の目的を伝えなければならない。しかしながら、クラウドソーシング・サービスにおいては、参加者の個人特定情報やE-Mailアドレス等を取得することが禁止されていることもあるため、ディブリーフィングの方法に注意する必要がある。このような時の対処法としては、タスク自体は完了したが、まだ結果が送信されていない段階で事後説明を行う方法が推奨されている（e.g. Office of Research Ethics, 2013）。また、参加者からの質問を受け付けるときも、参加者の連絡先を尋ねるのではなく、研究者の連絡先を事後説明文書等に記載しておくといった方法を用いる必要がある。

2.4.4 自由意志による参加と中断

参加者の自由意志による研究参加については、以下の点に配慮が必要である。受注者としての参加者は、発注者としての研究者から

タスク達成度を評価され、達成度が基準値以下である場合は、報酬を支払われないという状況に置かれている。このことが、参加者に対して言外の圧力となり、タスクの遂行途中での中断を抑制するように働くかも知れない。そのようなことを避けるためには、事前のインフォームド・コンセントの段階で、研究への参加をいつでも中断できること、中断しても一切不利益がないことを十分に説明し、また、それを保証する必要がある (Behrend et al., 2011; Office of Research Ethics, 2013)。

2.4.5 研究倫理審査

研究倫理について、さらに考慮すべき点は、研究倫理審査の問題である。行動研究を行う場合、研究者の所属機関、および研究が行われる組織の研究倫理委員会 (institutional review board, IRB, または, independent ethics committee, IEC) か、それに相当するものに、研究計画を提示した上で、承認を受けることが求められている。しかしながら、クラウドソーシング・サービスは、これまでにない形態の労働ポータルであり、研究計画の審査・承認を行う倫理審査委員会が、その種のサービスを用いた研究の実情についてよく理解していない可能性がある。従って、参加者の倫理面での扱いについては、研究者が十分に配慮し、また、先行する事例等を参考にしながら十分な説明を行っていく必要があるだろう。

3. 日本におけるクラウドソーシング・サンプルを用いたオンライン調査・実験環境の構築へ向けて

3.1 クラウドソーシング・サービスとオンラインのサーヴェイプラットフォームの併用

本稿では、行動研究におけるオンラインの

サーヴェイプラットフォームと、参加者プールとしてのクラウドソーシング・サービスの利用に関わる諸問題を、先行研究の成果に基づいて論じてきた。最後に、それらの内容をもう一度整理しながら、日本における同種の調査・実験環境の構築の際に考慮すべき点を挙げてみたい。

まず、クラウドソーシング・サービスとオンラインのサーヴェイプラットフォームを併用して行動研究を行うことの利点は、研究実施に要する時間・労力面でのコストを大幅に軽減できることにある。さらに、オンラインでの調査を行う場合、これまではそのような調査に特化したサービスを提供する専門の調査会社を使うことが多かったが、クラウドソーシング・サービスとサーヴェイプラットフォームを併用する方法では、専門の調査会社を利用する場合に比べて金銭的成本も低く抑えることができる。また、既にある各種のサポートツール (例えば, jsPsych) の中には、典型的な調査・実験で使うことができるようなプログラムコードを共有する仕組みが用意されているものもあり、それらを利用することで、研究準備に必要な時間をさらに削減することも可能である。また、オンライン上の調査・実験で得たデータは、従来型の実験室実験や大学生サンプルから得られたデータと比べて遜色ない程度の信頼性が保たれていることが、質問票ベースのデータのみならず、正確な時間制御を必要とする認知実験においても確かめられている。

一方で、参加者が同一の、あるいは類似の研究に重複参加することで生じる、参加者の非ナイーブ性の問題や、オンライン調査・実験に特有の脱落率の高さや、新しい技術であるが故の技術的ハードルの高さ、あるいは倫理面での配慮など、従来型の調査・実験では想定しなかったような新規の問題があることも事実である。

これらの問題は、研究者側の適切な配慮に

よって解決されるべきであり、そのために必要な労力が発生することが新たな問題として存在するのは事実であるが、クラウドソーシング・サービスとオンライン・サーヴェイプラットフォームの併用には、これらのデメリットを補ってあまりある利点が存在すると考えられる。

3.2 日本における導入にあたって考慮すべき問題

本稿では、主として Amazon Mechanical Turk というクラウドソーシング・サービスを参加者プールとして利用した先行研究に基づいた議論を行ってきた。その結論としては、このようなクラウドソーシング・サービスの導入は、行動研究において大きなメリットをもたらすと予想されることを指摘できる。しかしながら、本稿の執筆時点（2014年11月）では、AMTで Requester、すなわち発注者となるためには、米国の電子小切手決済対応の銀行口座と米国内の請求者住所を持っている必要がある⁹。したがって、日本国内からは、そのままではタスクを直接 AMT に対して送信することができない。日本国内の研究者が AMT にタスクを登録するためには、AMT との仲介を行う第三者を介してタスクを登録する必要がある¹⁰。一方で、AMT の Worker、すなわち受注者は、米国籍を有していなくても良く、多くの米国外の Worker が存在する。しかしながら、Behrend et al. (2011) によると、Worker の人種別分布では白人が大多数（82.2%）を占め、アジア系は白人に次いで登録者がいるものの、その数は極めて少なく（8.3%）、日本人を対象とした調査・実験を行う場合は、AMT 以外のサービスを通じて参加者をリクルートする必要がある。

日本国内では、ランサーズ株式会社 (<http://www.lancers.jp>)、株式会社クラウドワークス (<http://crowdworks.jp>) 等のクラウ

ドソーシング・サービスを提供する会社が複数存在しており（総務省, 2014が詳しい）、国内において研究目的でのクラウドソーシング・サービスを利用する場合は、そちらを利用することになるだろう。しかしながら、AMT とは異なり、クラウドソーシング・サンプルの人口統計変数上の特性や、行動研究サンプルとしての信頼性、妥当性については十分に検討がされているとは言えない。日本国内でのクラウドソーシング・サンプルを対象とした、行動研究サンプルとしての適切性について、データに基づいた検証を行う必要がある。

さらに、オンラインでの調査・実験を行うにあたって、どのような情報通信機器が用いられるかについての検討と、それに応じた調査・実験環境の整備も必要だろう。本稿で言及した先行研究では、多くの場合、参加者は、PC とそれにインストールされた Web ブラウザを用いて研究に参加していた。日本は、米国、英国、フランス等に比べて、情報通信機器の個人保有状況における PC、スマートフォン、タブレット等の普及率がやや低い一方で、いわゆるガラケー（フィーチャーフォン）の保有率が他国と比べて高いことが指摘されている（総務省, 2014）。参加者がどのような情報通信機器を介して、クラウドソーシング・サービスやオンラインのサーヴェイプラットフォームにアクセスしてくるかによって、特に認知実験等の実行可能性は変化する。例えば、認知実験をタスクとして登録する場合は PC の利用を必須とするなどの注意も必要になってくるであろう。

本稿での議論が、日本国内の研究者によるクラウドソーシング・サービスとサーヴェイプラットフォームの利用推進の一助となれば幸いである¹¹。

注

¹ 後述する Amazon Mechanical Turk (AMT) のサービス開始は2005年11月である。

- ² 2014年時点で190カ国以上、50万人以上の Worker が登録しているとされる (AMT の利用に伴う技術的諸問題については、Mason & Suri, 2012が詳しい)。
- ³ AMT の場合、決済はクレジットカードによって行われ、HITs への報酬額に上乘せされる10% のコミッションを Amazon に支払うようになっている。
- ⁴ 4～5種類の尺度を用いる調査型の HITs の場合、報酬の中央値は80セントという報告もある (Paolacci & Chandler, 2014)。
- ⁵ AMT では、各 Worker は WorkerID という固有の識別用 ID が発行されるため、この ID を重複参加禁止のフィルタとして用いることによって、同一 Worker によるデータの重複を防ぐことができる。
- ⁶ この調査は、AMT のサービス開始以前に実施されたものである。
- ⁷ jsPsych については、以下の Web サイトにドキュメントとダウンロードのためのリンクがある。<http://docs.jspsych.org>
- ⁸ そのようなツールとしては、psiTurk (<https://psiturk.org>) があるが、その操作はコマンド入力により行わなければならない、一定程度以上のコーディング能力が要求される。
- ⁹ <http://aws.amazon.com/jp/mturk/faqs/>
- ¹⁰ 例えば、そのような仲介を行う企業としては CrowdFlower (<http://www.crowdflower.com>) が知られている。
- ¹¹ 本稿は、先行研究の概観を元に、今後考える研究方法の変化について論じたものであるが、本稿の著者は、実際にはクラウドソーシング・サービス、および有償のサーヴェイプラットフォームを併用した調査・実験を行ったことがあるわけではない。既に利用された方からのコメントをいただければ幸いである。

引用文献

- Amir, O., Rand, D. G., & Gal, Y. a. K. (2012). Economic games on the Internet: The effect of \$1 stakes. *PLoS ONE*, 7, e31461.
- Behrend, T., Sharek, D., Meade, A., & Wiebe, E. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, 43, 800–813.
- Birnbaum, M. H. (2000). *Psychological experiments on the Internet*. San Diego, CA: Academic Press.
- Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology*, 55, 803–832.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon’s Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6, 3–5.
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, 46, 112–130.
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon’s Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*, 8, 1–18.
- de Leeuw, J. (2014). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 1–12. [Advance Online Publication.]
- Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, 26, 213–224.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior Research Methods*, 44, 1–23.
- 日本心理学会 (2009). 社団法人日本心理学会倫理規程. 東京: 社団法人日本心理学会.
- O’Neil, K., & Penrod, S. (2001). Methodological variables in Web-based research that may affect results: Sample type, monetary incentives, and personal information. *Behavior Research Methods, Instruments, & Computers*, 33, 226–233.
- Office of Research Ethics. (2013). *Human Participant research guidelines – Use of crowdsourcing services*. Waterloo, ON, Canada: University of Waterloo. Retrieved Nov, 3, 2014, from <https://uwaterloo.ca/research/>

- sites/ca_research/files/uploads/files/crowd_sourcing_guidelines_access_check_done.pdf.
- Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23, 184-188.
- Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5, 411-419.
- Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, 49, 243-256.
- Schnoebelen, T., & Kuperman, V. (2010). Using Amazon Mechanical Turk for linguistic research. *Psihologija*, 43, 441-464.
- 総務省 (2014). 平成26年版 情報通信白書. 東京: 総務省. Retrieved Oct. 29, 2014, from <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h26/pdf/26honpen.pdf>
- Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods*, 43, 155-167.
- Stanton, J. M., & Rogelberg, S. G. (2001). Using Internet/Intranet Web pages to collect organizational research data. *Organizational Research Methods*, 4, 200-217.
- Suri, S., & Watts, D. J. (2011). Cooperation and contagion in Web-based, networked public goods experiments. *PLoS ONE*, 6, e16836.
- Terdiman, D. (2005). Amazon looks to solve problems that stump computers. *CNET News*. Retrieved Oct. 30, 2014, from http://news.cnet.com/2100-1007_3-5942800.html