序　文

本書のアイデアが生まれたのは，2005 年，コロンビア大学の地下室だった。その当時，大学院生だった私は，後で博士論文となるオンライン実験を行っている最中だった。この実験の科学としての部分は第 **4** 章で説明しよう。ここでは博士論文そのほか，どの論文でも触れなかった部分について話したい。それこそが私の調査に対する考えを変えたのだ。ある朝，地下のオフィスに来てみると，一夜にしてブラジルから 100 人もの人が私の実験に参加してくれたことがわかったのだ。このシンプルな経験は私にとって大変な衝撃だった。当時，私の友人が従来型の実験室実験を行っていたので，実験参加者を募り，監督し，手当を支払うことがどれだけ大変なことかはわかっていた。この場合，1 日で 10 人も集めることができれば，よい進捗だといえるだろう。だが，私のオンライン実験では，**私が寝ている間に**，100 人が参加してくれたのだ。寝ている間に研究が進むなんて夢のような話は信じられないかもしれない。だが本当なのだ。技術の変化——とりわけアナログ時代からデジタル時代への移行——によって，私たちはいまやソーシャルデータを新たな仕方で集め，分析することができるのだ。本書は，このように新しい方法で社会調査を行うための本である。

本書はもっとデータサイエンスをしたい社会科学者と，もっと社会科学をしたいデータサイエンティスト，そしてこの 2 つの学問分野のハイブリッドに興味をもつすべての人のための本だ。そうだとすれば，この本が単に学生と大学教員のための本ではないということはいうまでもないだろう。私は今でこそ大学（プリンストン大学）で働いているが，以前は政府（米国国勢調査局）やテック産業（Microsoft Research）で働いたこともある。大学の外でたくさんのすばらしい研究が行われていることも知っている。だから，どこで働いていても，どんな技術を今使っているとしても，もしあなたが自分のしていることが社会調査だと思うなら，本書はあなたのために書かれた本だ。

もう気づいているかもしれないが，本書のトーンはその他のたくさんの学術書とは少し違う。意図的にそうしているのだ。本書は 2007 年からプリンストン大学の社会学部で教えている計算社会科学の大学院向けゼミから生まれた。ゼミの興奮やエネルギーを本書でもある程度伝えたいと思っている。具体的にいえば，

本書は3つの特徴をもつよう心がけている。役に立つこと，未来志向であること，楽天的であること。

役に立つ： 私の目標は読者の役に立つ本を書くことだ。だから，オープンで形式ばらない，そして事例を使って説明するスタイルを採用したい。私の伝えたいもっとも大事なことは，社会調査についてのある種の思考の作法だ。そして，経験上，この種の思考法を伝える最善の方法は，形式ばらないスタイルで叙述すること，そして多くの事例を用いることだと考えている。さらに，各章末に，「読書案内」を設けて，紹介したトピックの多くについてより詳細で専門的な本に進めるよう手助けをしている。最終的には，自分自身で研究すること，そして他人の研究を評価すること，この両方に本書が役立てばよいと願っている。

未来志向： 本書が今日すでに存在するデジタルシステムだけでなく，将来生み出されるだろうデジタルシステムを使った社会調査に役立つようにと考えている。私がこの種の研究を始めたのは2004年のことだが，それ以来，多くの変化があったし，あなたのキャリアが進むにつれてさらに多くの変化が生じるに違いない。変化する状況下でも時代遅れにならないためのこつは抽象化だ。たとえば，本書では，現在のTwitter APIの使い方を説明することはしない。そうではなく，ビッグデータから何を得ることができるかを伝えたい（第2章）。また，Amazon Mechanical Turk[1)]で実験をするための手順を逐一教えるわけでもない。そのかわり，デジタル時代のインフラを用いた実験の設計と解釈の方法を教えたい（第4章）。抽象化することで，タイムリーなトピックについての古びることのない本になれば，と考えている。

楽天的： 本書が対象とする2つのコミュニティ――社会科学者とデータサイエンティスト――はその学問的背景も興味関心もずいぶん異なる。本書では科学としての両者の違いについて論じるつもりだが，それに加えて，2つのコミュニティはスタイルの点でも違っている。データサイエンティストは一般的に楽天的だ。彼らはコップに水が半分も入っていると考える傾向がある。他方で，社会科学者は一般的により批判的だ。彼らはコップに水が半分しか入っていないと考える傾向がある。本書は，データサイエンティストの楽天的トーンを採用したい。だから，例を紹介するときには，私がなぜこの例が好きなのかを述べるようにするつもりだ。そして研究の問題点を指摘するときには――どんな研究も完璧

訳注●
1） クラウド上でマイクロタスクを依頼することのできるクラウドソーシングサービス。研究目的では，サーベイや実験の参加者を募るために用いられる。

ではありえないからそうすることもある――，問題をポジティブに，楽天的なやり方で指摘するように努めよう。批判のための批判はしたくない――批判するのは，そうすることであなたがよりよい研究をするために役立つからだ。

デジタル時代の社会調査が始まってまだ日が浅い。しかし，ある種の誤解がまん延しているようなので，この序文で，これを正しておくのも無駄ではないだろう。データサイエンティストの側では，2つのよくある誤解をみかける。1つは，より多くのデータがあれば問題が自動的に解決されるはずだと考える思考法だ。しかし，私の経験に基づく限り，社会調査についてこれはあてはまらない。実際，社会調査にとっては，よりよいデータ――より多いデータではなく――の方が役に立つといってよいだろう。データサイエンティストから聞こえてくる第2の誤解は，社会科学とは常識を大げさに飾り立てただけの話法の集まりだという考えだ。もちろん，社会科学者として――もっといえば社会学者として――これに同意するわけにはいかない。頭のよい人たちが，人間行動を理解するために，長い間，懸命に努力してきたのだ。そうして積み上げられた知識を無視するのは賢いやり方ではない。私は，本書がそうした知識をわかりやすく提供するものになるはずだと思いたい。

社会科学の側にもよくある誤解が2つある。第1に，たまたまごく少数の不出来な論文を読んでしまったことで，デジタル時代のツールを用いた社会調査という発想自体を否定する人たちがいる。本書を手にとっている以上，あなたもおそらく，陳腐な，または間違った（あるいはその両方）仕方でソーシャルメディアデータを使った論文を，すでにいくつか読んだことがあるかもしれない。私も読んだことがある。しかし，こうした例があるからといって，デジタル時代の社会調査がすべてダメだと決めつけるのは深刻な間違いだ。実際，あなたは陳腐なもしくは間違った仕方でサーベイデータを使った論文だって読んだことがあるだろう。しかし，だからといって，サーベイを用いた研究そのものを否定することはないはずだ。サーベイデータを用いた優れた研究があることを知っているからだ。本書では，デジタル時代のツールを用いた優れた研究もまた存在するということを示したいと思う。

社会科学者から聞くよくある誤解の2番目は，現在と将来を取り違えることだ。私がこれから説明しようと考えるデジタル時代の社会調査について検討する場合，2つの異なる問いを区別することが重要だ。「このスタイルの調査が現時点でどのくらいうまくいくのか」という問いと，「このスタイルの調査が将来どのくらいうまくいく可能性があるのか」という問いだ。研究者は第1の問いに

答えるべく訓練されているものだが，本書では第2の問いの方がより重要だと考える。つまり，デジタル時代の社会調査がまだパラダイムを変えるような巨大な知的貢献をなしえていないとしても，デジタル時代の調査は急速に進歩しているのだ。この，現在のレベルを超えていく変化のスピードこそ，デジタル時代の調査を私にとって刺激的なものとしているといえる。

いつかわからない未来のどこかの儲け話をしているかのように聞こえるかもしれない。だが，何か特定の研究を売り込むつもりはない。私は，Twitter，Facebook，Google，Microsoft，Appleその他のテック企業の株をもっているわけではない（ただ完全な情報公開をしておくと，Microsoft，Google，Facebookで働いたり，研究資金を受け取ったりしたことはある）。だから本書の目標は，信頼できる語り手として，今可能となっている刺激的な新しい研究のすべてを伝えること，そして陥りがちな（私自身もときおり陥った）いくつかの罠を避けるよう手助けすることだ。

社会科学とデータサイエンスの融合は，ときに計算社会科学と呼ばれる。これをテクニカルな分野と考える人もいるが，本書は伝統的な意味での技術書ではない。たとえば，本文には数式は1つも出てこない。私がこの形式で書くことを選んだ理由は，ビッグデータ，サーベイ，実験，マスコラボレーション，倫理を含むデジタル時代の社会調査についての包括的な展望を描きたかったからだ。これらのトピックをすべてカバーしつつ，それぞれのテクニカルな詳細を説明することは不可能だということがわかった。そこで，より専門的な内容については，各章末の「読書案内」と「数学ノート」の箇所で示唆しておくことにした。言い換えると，本書は，何か特殊な計算の方法を教えるための本ではなく，社会調査についての思考法を変えるために作られた本なのだ。

本書の授業での使い方

先に述べたように，本書の一部は2007年以来プリンストン大学で教えている計算社会科学の大学院ゼミから生まれた。本書を使って授業をしようと考えている人がいるならば，本書が私の授業からどうやって生まれたか，他の授業でどうやって使われることを想定しているかを説明するのが役立つだろう。

何年間も私は教科書なしで授業をしていた。いろいろな論文を読書課題として与えていただけだった。学生は論文を読んで学習することはできたが，論文だけでは，私が思い描いていたような思考法自体の変化を引き起こすことはできなかった。だから，私は学生が大局的な見方ができるように，授業で視点や文脈，助

言を与えることに大半の時間を費やすことになった。本書は，この視点，文脈，助言を，社会科学やデータサイエンスの予備知識無しでも消化できるように書き下す試みである。

セメスター単位の授業では，本書とさまざまな追加文献を組み合わせることをおすすめする。たとえば，実験の章に 2 週費やすとして，第 4 章といっしょに，実験計画・分析における処置前情報の役割，企業の大規模 A/B テストにまつわる統計上・計算上の問題，メカニズムの分析に特化した実験計画，Amazon Mechanical Turk などのオンライン労働市場からの参加者を用いることにともなう実践的・科学的・倫理的問題といったトピックについての読書課題を与えるのもよいだろう。プログラミングに関する読書課題や演習を組み合わせることもできる。どの組み合わせがよいかは授業を履修する学生が誰か（学部生か修士課程か博士課程か），彼らのバックグラウンドや目的によるだろう。

セメスター単位の授業で週ごとに課題を課すこともできる。各章には難易度ごとにラベル付けした各種の課題がついている。やさしい（ ），普通（），難しい（），とても難しい（），だ。問題には必要スキルに関するラベルも付けてある。要数学（ ），要コーディング（），データ収集（）のラベルだ。さらにいくつかの課題には私のお気に入りというラベルもつけている（）。さまざまな課題を用意したので，あなたの学生に適切な課題をみつけることができるはずだ。

本書を授業で使う際に役立つように，各章ごとのシラバスやスライドなどの講義資料，各章ごとのおすすめ文献，いくつかの課題に対する解答を用意し始めている。これらの資料は http://www.bitbybitbook.com/ に用意した。追加すべき資料があれば情報提供をお願いしたい。

目　次

序　文　iii

第1章　イントロダクション　1
1.1　ロールシャッハテスト　1
1.2　デジタル時代へようこそ　2
1.3　研究デザイン　5
1.4　本書のテーマ　6
1.5　本書のアウトライン　9
読書案内　11

第2章　行動を観察する　13
2.1　イントロダクション　13
2.2　ビッグデータ　14
2.3　ビッグデータに関する10の特徴　17
2.3.1　巨大さ　18
2.3.2　常時オン　21
2.3.3　非反応性　24
2.3.4　不完全性　25
2.3.5　アクセス不能性　27
2.3.6　非代表性　30
2.3.7　ドリフト　34
2.3.8　アルゴリズムによる交絡　36
2.3.9　汚　染　38
2.3.10　センシティブ　41
2.4　研究戦略　42
2.4.1　事物を数える　43
2.4.2　将来予測と現在予測　48

2.4.3 実験に近づける　52
2.5 結　論 …… 62
数学ノート　64　読書案内　71　課題　79

第3章 質問をする — 87
3.1 イントロダクション …… 87
3.2 質問すること vs. 観察すること …… 90
3.3 総調査誤差フレームワーク …… 91
3.3.1 代 表 性　93
3.3.2 測　定　96
3.3.3 コ ス ト　100
3.4 誰に質問するか …… 101
3.5 新しい質問の仕方 …… 109
3.5.1 生態学的経時的評価法　111
3.5.2 ウィキサーベイ　114
3.5.3 ゲーミフィケーション　117
3.6 ビッグデータにサーベイを結びつける …… 120
3.6.1 測定項目増加法　120
3.6.2 調査対象拡張法　125
3.7 結　論 …… 133
数学ノート　133　読書案内　138　課題　143

第4章 実験を行う — 149
4.1 イントロダクション …… 149
4.2 実験とは何か？ …… 151
4.3 実験の 2 つの次元：ラボ／フィールドとアナログ／デジタル … 154
4.4 シンプルな実験を超えて …… 160
4.4.1 妥 当 性　164
4.4.2 処置効果の不均質性　169
4.4.3 メカニズム　172
4.5 自分でやってみる …… 176
4.5.1 すでにある環境を利用する　176

4.5.2 自分自身で実験環境を構築する 181
4.5.3 自分自身で製品を開発する 185
4.5.4 頼りになるパートナーと組む 186
4.6 アドバイス …… 191
4.6.1 変動費用ゼロのデータを作る 192
4.6.2 実験デザインに倫理を組み込む：置き換える（replace）・洗練させる（refine）・減らす（reduce） 198
4.7 結　論 …… 204
数学ノート 205　読書案内 210　課題 222

第5章 マスコラボレーションを生み出す —— 233
5.1 イントロダクション …… 233
5.2 ヒューマンコンピュテーション …… 236
5.2.1 ギャラクシー・ズー 237
5.2.2 政治的マニフェストのクラウドコーディング 245
5.2.3 結　論 248
5.3 オープンコール …… 250
5.3.1 Netflix プライズ 251
5.3.2 フォールディット 254
5.3.3 ピアトゥパテント 257
5.3.4 結　論 260
5.4 分散データ収集 …… 262
5.4.1 eBird 263
5.4.2 フォトシティ 265
5.4.3 結　論 269
5.5 あなた自身のものをデザインする …… 272
5.5.1 参加者を動機づける 272
5.5.2 不均質性を利用する 273
5.5.3 集中させる 274
5.5.4 驚きを可能にする 275
5.5.5 倫理的であること 276
5.5.6 デザインに関する最後のアドバイス 278
5.6 結　論 …… 280

読書案内　281　　課題　287

第6章　倫　　理 ——— 291

6.1　イントロダクション …… 291
6.2　3つの事例 …… 294
6.2.1　感情の伝染　294
6.2.2　嗜好，紐帯，時間　295
6.2.3　アンコール　296
6.3　デジタルはこれまでと違う …… 298
6.4　4つの原理 …… 302
6.4.1　人格の尊重　303
6.4.2　善　　行　304
6.4.3　正　　義　307
6.4.4　法と公益の尊重　308
6.5　2つの倫理枠組み …… 310
6.6　特に難しい領域 …… 312
6.6.1　インフォームドコンセント　312
6.6.2　情報リスクの理解と管理　315
6.6.3　プライバシー　323
6.6.4　不確実な状況での意思決定　326
6.7　実践のための秘訣 …… 329
6.7.1　IRB はゴールではなくスタートライン　329
6.7.2　関係する人全員の立場に立ってみよ　331
6.7.3　研究倫理は 0 か 1 かではなく連続的に考えよ　332
6.8　結　　論 …… 332
歴史についての付録 …… 333
読書案内　338　　課題　347

第7章　未　　来 ——— 363

7.1　先 を み る …… 363
7.2　未来のテーマ …… 363
7.2.1　レディメイドとカスタムメイドのブレンド　363

7.2.2 参加者中心のデータ収集 364
7.2.3 研究デザインにおける倫理 366
7.3 はじまりに戻って …… 367

謝　辞 369
訳者あとがき 375
参考文献 379
索　引 440
著者紹介・訳者紹介 454

訳文中の（ ）は原書に準じている。[ ] は訳者が補ったものである。また，原書でイタリック体で強調されている語は，訳文中ではゴシック文字で表示した。

本書のコピー，スキャン，デジタル化等の無断複製は著作権法上での例外を除き禁じられています。本書を代行業者等の第三者に依頼してスキャンやデジタル化することは，たとえ個人や家庭内での利用でも著作権法違反です。

第 1 章
イントロダクション

1.1 ロールシャッハテスト

2009 年の夏，ルワンダ中の携帯電話が鳴っていた。家族，友人，仕事関係者からの数百万の呼び出しに加えて，約 1000 人のルワンダ人が Joshua Blumenstock からの呼び出しを受けていたのだ。この研究者たちはルワンダの富と貧困について研究するため，ルワンダ最大の携帯電話会社の 150 万の利用者のデータベースに基づくランダムサンプルサーベイを行っていた。Blumenstock たちはランダムに選択された人々に，サーベイに参加してもよいかを聞き，研究の趣旨を説明し，人口学的，社会的，経済的特徴に関する一連の質問項目を尋ねた。

ここまでの説明を聞いただけなら，従来の社会科学のサーベイと変わらないじゃないか，と思われるかもしれない。だが，ここから先が従来とは違うのだ——少なくとも現時点では。Blumenstock たちは，サーベイデータに加えて，150 万人の完全な通話記録を集めたのだ。彼らはこの 2 つのデータソースを組み合わせた。まず，サーベイデータを使って，通話記録から人々の富を予測可能な機械学習モデルを訓練する。その後，このモデルを用いて，データベース上の 150 万人の携帯電話利用者の富を推定するのだ。彼らはさらに通話記録に埋め込まれた地理情報を用いて 150 万人の利用者の居住地も推定した。これら，推定された富と推定された居住地を合わせることにより，ルワンダにおける富の地理分布について高解像度の地図を作成することが可能となる。より詳しくいえば，ルワンダの最小の行政単位である 2148 セルのそれぞれの富を推定したのだ。

かつて，ルワンダのここまで細かい地理的区画に関して，富の推定が行われたことはない。だから，この推定値の妥当性を検証することはできない。ただ

Blumenstock たちが推定値を集計してルワンダの 30 の地域にまとめてみると，その推定値は，途上国のサーベイの参照基準と広く認められている『人口保健調査』（Demographic and Health Surveys）の推定値にきわめて近いことがわかった。この場合，2 つの方法が生み出す推定値は似ていたが，Blumenstock たちの方法は伝統的な『人口保健調査』よりも約 10 倍速く約 50 倍安価に行うことができた。この劇的なほど迅速で安価な推定法を用いることで，研究者，政府，企業は新たな可能性を手にすることになる（Blumenstock, Cademuro, and On 2015）。

この研究は，図柄が何にみえるかはみる人の背景によるという点で，ロールシャッハのインクブロットテスト[1)]に似ている。**社会科学者**の多くはここに，経済発展理論を検証するために利用可能な新たな測定ツールをみるだろう。**データサイエンティスト**の多くは，クールな新しい機械学習の問題をみるだろう。**ビジネスピープル**の多くは，自分たちのすでに集めているビッグデータの潜在的価値を活かすための強力な方法をみるだろう。**プライバシー運動家**の多くは私たちが大衆監視の時代に生きていることの不気味な証左をみるだろう。そして最後に，**政策立案者**の多くは新技術がよりよい世界を生み出すために役に立ちうるということをみるだろう。実際，この研究は今述べたすべての側面をもっている。このようなさまざまな側面を併せ持つ研究だからこそ，私はこの研究を社会調査の未来を映す窓としてみているのだ。

1.2　デジタル時代へようこそ

デジタル時代はいたるところに現れ，存在感を増しつつあり，研究者のできることを変えていく。

本書の大前提は，デジタル時代は社会調査の新たな可能性を切り開くはずだという点にある。研究者は今，最近までまったく不可能であったやり方で，行動を観察し，質問をし，実験を行い，コラボレーションすることができる。こうした新たな機会とともに新たなリスクも生じている。研究者はいまや，近年まではありえなかった形で，人々に危害を与えてしまうかもしれない。この機会とリスクが生み出されたのは，時代がアナログからデジタルへ移りつつあるからだ。この移行は――あたかも電灯をつけるがごとく――一気に生じるわけではない。実際

訳　注
1）　紙にたらしたインクの図形が何にみえるかを答えさせ，性格を判断しようとする心理テスト。

のところ，まだ完全に移行が完了したわけではない。だが，何か大きな変化が進行しつつあるということは今でも十分にみてとれるはずだ。

この移行を実感するには，日常生活の変化を探してみるのが 1 つの方法だ。かつてアナログだった多くの生活上のものごとがいまやデジタルになっている。もしかしたらあなたはかつてフィルムカメラを使っていたかもしれないが，今はデジタルカメラ（おそらくスマートフォンの一部の）を利用しているはずだ。かつては，紙の新聞を読んでいたかもしれないが，今はオンラインの新聞を読んでいるだろう。かつては現金払いをしていたかもしれないが，今はクレジットカードを使っているはずだ。どの場合でも，アナログからデジタルへ移行するということは，あなたに関するデータがますます多く，デジタルに取得され保存されるということだ。

事実，全体としてみれば，移行の帰結は驚くべきものだ。世界の情報量は急速に増えつつあり，ますます多くの情報がデジタルに保存され，分析，伝達，統合が可能になっている（図 **1.1**）。このデジタル情報の全体が「ビッグデータ」と呼ばれるようになっている。デジタルデータの爆発に加えて，計算能力へのアクセスに関しても同じような発展がみられる（図 **1.1**）。このトレンド——デジタルデータの増大と計算能力の利用の増加——は当面続く可能性が高い。

社会調査という目的にとって，デジタル時代のもっとも重要な特徴は，**いたるところにコンピュータ**が存在するということにあると思う。政府や大企業だけが利用できた，部屋の大きさほどのマシンから始まって，コンピュータのサイズはどんどん小さくなり，どこにでも存在するようになっていった。1980 年代以降の 10 年ごとに新たなコンピュータが出現してきた。パーソナルコンピュータ，ラップトップ，スマートフォン，「モノのインターネット」（たとえば，車や時計，サーモスタット等のデバイス内部のコンピュータ）に埋め込まれたプロセッサ（Waldrop 2016）。こうした遍在するコンピュータは単に計算するだけでなく，情報を取得し，保存し，伝達するようになりつつある。

研究者にとっては，コンピュータがいたるところに存在することで何が生じているのかを確認する一番簡単な方法は，オンラインをみることだろう。オンラインとは，完全に測定でき実験も可能な環境のことでもある。たとえば，オンラインストアでは，数百万の顧客の購買パターンについて信じがたいほど正確なデータを簡単に収集できる。さらに，顧客のグループをランダム化して異なるショッピング体験をさせることも可能だ。購買履歴の追跡に加えてランダム化もでき

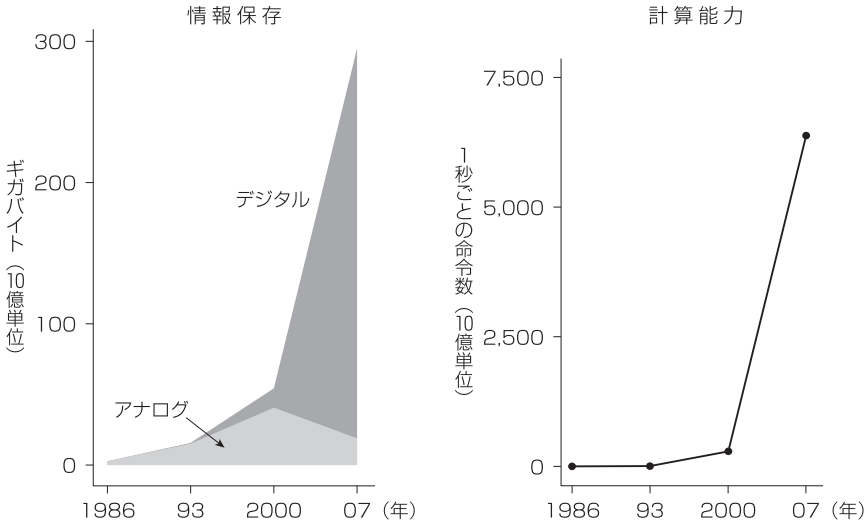

図 1.1　情報保存容量と計算能力の劇的な増加　これらの増大に加えて，情報保存はいまやほぼ完全にデジタルになっている。この変化が社会調査の研究者に信じがたいほどの機会を生み出しているのだ。Hilbert and López（2011）の図 2，図 5 より。

る。ということはつまり，オンラインストアでは日々ランダム化統制実験[2)]を実施することができるということだ。実際，もしあなたがオンラインストアから何か買ったことがあるならば，行動が記録されているだけでなく，あなたが気づいていようといまいと，ほぼ確実に何らかの実験に参加していることになるのだ。

この完全に測定され，完全にランダム化可能な世界は単にオンラインでだけ起きているのではなく，ますますいたるところに現れつつある。物理的店舗はすでにきわめて詳細な購買データを集めているし，顧客の購買行動をモニタし，ビジネスのルーチンに実験を織り交ぜつつある。「モノのインターネット（Internet of Things）」とは，物理的世界の行動がデジタルセンサーによりますます補足されるようになるということだ。言い換えると，デジタル時代の社会調査について考えるためには，**オンライン**だけでなく，**あらゆるところで**デジタル化が生じていると考えるべきなのだ。

行動の測定と処置のランダム化に加えて，デジタル時代は人々がコミュニケートするための新たな方法も生み出している。新たなコミュニケーション形式により，研究者は革新的なサーベイを走らせることも，同僚や一般市民とマスコラボ

訳　注
2）　本書第 4 章を参照。

レーションを作り出すこともできるのだ。

疑い深い人の中には，こうした性能は何も新しいものではないと指摘する人もいるかもしれない。というのは，過去にも，コミュニケーション能力を高める大きな技術革新（たとえば，電報〔Gleick 2011=2013〕）が存在したし，コンピュータは 1960 年代からほぼ一定の速度で計算速度を増しているから，というわけだ（Waldrop 2016）。だがこうした疑い深い人が見落としているのは，同じものが多く集まれば，ある時点で別のものになるということだ（Halevy, Norvig, and Pereira 2009）。私の好きなアナロジーはこうだ。馬の 1 枚のイメージを写せば，それは写真である。だがもし 1 秒に 24 個の馬のイメージを写せば，それは映画になる。もちろん，映画とは単なる写真の集まりではある。だが，よほど頑固な懐疑論者でなければ，写真と映画が同じものだとはいわないだろう。

研究者は今，写真撮影法から映画撮影法への移行と同じような変化の過程にいる。とはいえ，この変化によって，過去に学んだすべてのことは忘れ去られるべきだ，ということにはならない。写真撮影法の原理が映画撮影法を学ぶにも有益なように，過去 100 年間発展してきた社会調査の原理は，今後 100 年間の社会調査について考えるためにも有効だ。だが，変化によって，同じことをただ続けさえすればよい，とはいかなくなったのも事実だ。そうではなく，過去のアプローチを，現在と未来になしうることと組み合わせなければならないのだ。たとえば，Joshua Blumenstock らの調査は伝統的なサーベイ調査といわゆるデータサイエンスとの融合だ。どちらの要素も必要である。サーベイの解答だけでも，通話記録だけでも，貧困についての高解像度の推定は不可能だ。より一般的にいえば，社会調査の研究者は，デジタル時代の生み出した機会を活用すべく社会科学のアイデアとデータサイエンスのアイデアを組み合わせる必要があるだろう。どちらのアプローチもそれだけでは十分ではないのだ。

1.3 研究デザイン

研究デザインとは問いと答えを結びつけるデザインである。

本書はお互いから多くを学ぶことのできる 2 組のオーディエンスを対象として書かれている。一方は社会科学者で，社会行動を研究する訓練と経験を積んでいるが，デジタル時代の生み出した機会を利用することには慣れていない。他方の研究者たちはデジタル時代のツールを使うことは得意である一方，社会行動の

研究については新参者だ。この第2のグループを名づけるのは難しいが，ここではデータサイエンティストと呼んでおこう。データサイエンティストたち――多くの場合，コンピュータサイエンス，統計学，情報学，工学，物理学のような分野で訓練を積んでいる――は，必要となるデータや計算技術をもっているということもあり，デジタル時代の社会調査を一番最初に取り入れた人たちだった。本書は，この2つのコミュニティを1つにまとめあげることで，それぞれのコミュニティが単独で行うよりも実り多くかつ興味深い研究を可能にしようとする試みである。

この強力なハイブリッドを生み出すための最善の方法は，抽象的社会理論や高度な機械学習をことさらに利用することにはない。最善の出発点は，**研究デザイン**にある。社会調査を人間行動に対する問いかけとその答えからなる過程と考えるならば，研究デザインは結合組織（connective tissue）のようなものだ。研究デザインは問いと答えを結びつける。この結びつけを正しく行うことが説得力ある研究を行うための鍵となる。本書はあなたが過去に見聞きしたことのある――もしかしたら使ったことのある――4つのアプローチに焦点をあてる。すなわち，行動を観察すること，質問をすること，実験をすること，他者とコラボレーションすること。だが，新しいのは，デジタル時代になって，データを収集分析するためのこれまでとは異なる機会が開けたことだ。この新たな機会を用いて，これら古典的なアプローチを――取り替えるのではなく――現代的にすることが必要とされるのである。

1.4　本書のテーマ

本書の2つのテーマは①レディメイドとカスタムメイドの組み合わせ，そして②倫理だ。

2つのテーマが本書全体を特徴づけている。そこで，これらのテーマを今ここで強調しておけば，以後繰り返し登場するたび，それと認めることができるだろう。第1のテーマは，2人の偉人の比較を用いたアナロジーで説明しよう。マルセル・デュシャンとミケランジェロだ。デュシャンは『泉』などのレディメイドで多くの人に知られている。日常品を取り上げて芸術として転用（repurpose）したのだ。他方でミケランジェロは転用をしなかった。彼がダビデ像の製作を決心したとき，ダビデのようにみえる大理石を探すなどということはしなかった。彼

訳者あとがき

この本は，Matthew J. Salganik，2017，*Bit by Bit*，Princeton University Press の全訳である。

ビッグデータやデジタル実験から得られるデータの活用，ウェブやウェアラブルデバイスを用いた新しいサーベイ法，機械学習・自然言語処理・人工知能を用いた分析手法の援用など，デジタル時代の社会科学――これらを総称して「計算社会科学」と呼ぶことにしよう――は革命的な変化を遂げつつある。この本は，近年，世界的に発展しつつあるこの計算社会科学について，第一人者が単独で書き下ろした世界初の教科書だ。

著者の Matthew J. Salganik は 1976 年生まれ。米国のコロンビア大学社会学部にて博士号を取得した。指導教官は，ネットワーク科学の立役者で計算社会科学も牽引するあの Duncan Watts である。Salganik は本書でも紹介されているミュージック・ラボの実験で名を馳せ，その後も独創的な業績を多数あげている。現在はプリンストン大学社会学部教授として，計算社会科学をリードする中心人物の 1 人だ。

この本の内容自体は，きわめて平易な言葉で書かれており，あらためてここで解説するには及ばないだろう。社会科学やデータサイエンスの予備知識も必要ない。この本を通じて，読者は最先端のデータ分析や社会調査の理論と方法に触れ，学ぶことができるはずだ。

この本の対象は，社会科学やデータサイエンスの研究者に限られない。実務でデータ分析に従事している人，ビッグデータやソーシャルメディアが社会に与えるインパクトに関心をもつ人など，デジタル時代の社会調査に関心をよせるすべての人に開かれている。

ここで，訳者の考えるこの本の特徴をいくつか挙げよう。

第 1 に，この本の特徴は，データサイエンス的な先端性と社会科学的な正統性を両立させていることにある。この本には，先端的な手法やデータを用いた研究が多数登場するが，それらはすべて社会科学の伝統的な問題関心や理論枠組みの中に位置づけられ解説されている。だから，社会科学にバックグラウンドのある読者は，現在，ビッグデータや人工知能を援用して行われている最先端の研究

が，決して一過性の流行ではなく，社会科学の伝統的な問題に対する新たな取り組みとして必然性をもって発展してきたことが理解できる。一方で，データサイエンスになじんだ読者は，自分たちが普段行っている分析の社会科学的な意味付けを深く知ることができるだろう。著者がいうように，この本の価値は，研究の最前線を扱いながらも古びることのない，普遍的な社会科学上の問題を追究しているという点にある。

第 2 の特徴は，たくさんの研究事例を用いて，重要なポイントを解説していることだ。本文だけでなく，章末の読書案内や課題を含めれば，この本だけで計算社会科学の研究事例のかなりの部分をカバーしていることになる。読者はこれらの豊富な事例を通じて，最前線の社会科学・社会調査の全貌を学ぶことができる。そのためにも，読書案内や課題にも目を通してみてほしい。特に課題は，一読して頭の中で考えてみるだけでも理解度がかなり深まると思う（もちろん自分で手を動かして解いてみるのが一番だ）。

第 3 に，この本ではデジタル時代の社会調査にともなう倫理的問題について，深く扱っている。ビッグデータやデジタル実験などの先進的な研究領域には，未解決の倫理的問題がつきまとう。Facebook データを用いた世界各国の選挙への介入が疑われたケンブリッジアナリティカのスキャンダルは記憶に新しいが，デジタル社会調査によって研究者はかつてないほど，政治や人々の生活・プライバシーに介入する権力を手に入れている。Salganik はこのような権力には責任がともなうとし，倫理的問題を深く考察するための原理的な提言をしている。研究や実務で直面する倫理的問題と向き合うための助けとなるだろう。

この本の翻訳作業がスタートしたのは，2018 年の 1 月である。訳者の 1 人（瀧川）が原著の出版（2017 年 11 月）を知って，有斐閣の松井智恵子さんに翻訳の提案をしたのがきっかけだ。原著出版直後に翻訳をするのは，有斐閣としては異例のことだそうだが，松井さんの尽力により，企画を通すことができた。また，企画を通す過程では，一橋大学名誉教授の西口敏宏先生の強い推薦があったときく。記して感謝したい。

原著が出版されたのが 2017 年 11 月だから，それから 2 カ月足らずで翻訳を始めることができたことになる。この本の意義は古びることはないものの，やはり最先端の調査方法論をいち早く日本語圏の読者に伝えたいという思いから，2019 年の 4 月刊行を目標にして，瀧川が，法社会学，計算社会科学，社会調査等に造詣が深い常松，阪本，大林に声をかけ，翻訳チームを組み，共同で集中的

に翻訳に取り組むことにした。各自担当章を分担して翻訳したあと，すべての訳者でそれぞれの担当者の下訳を検討するため，有斐閣の会議室にて各章ごとの訳文検討会議を開いた。夏休みには仙台で合宿を開き，複数章の訳文の検討会を行ったのもよい思い出だ。

この本を日本語に翻訳するにあたっては，原著の平易で形式ばらない雰囲気をできるだけそのまま伝えるよう努力した。訳文は訳者全体のみならず外部の人の目も借りて検討したが，誤訳や読みにくい箇所も残っていることだろう。気づいた点があれば，ぜひ指摘してほしい。

この本の翻訳の過程で多くの人の協力を得た。最初に，異例に早い段階での翻訳の提案にもかかわらず，快く引き受けてくれて企画を通し，以後もさまざまな面で翻訳をサポートしていただいた松井さんに感謝したい。また，西口先生の推薦は訳者にとっても勇気づけられるものだった。校正段階では，校正者の茅（瀧川）しのぶさんに多数の誤りを直していただいた。翻訳に際しては，特に第5章の天文学関係の用語について，東北大学学際科学フロンティア研究所の當真賢二准教授，同研究所の市川幸平助教から貴重な示唆を得た。さらに，訳文全体は，東京大学大学院総合文化研究科の計算社会科学ゼミ（2018年度「世界システム1」「人間の安全保障演習V」）に参加した大学院生にも目を通してもらった。貴重なフィードバックを通して訳文の改善に貢献してくれた伊藤ジェームズ信玄さん，魚住知広さん，小菅生草太さん，小山達也さん，榛葉絵美さん，ディーン・リーさん，中村雄人さん，成田龍一朗さん，林稜さんに感謝申し上げたい。また訳者の1人である大林は，青山学院大学社会情報学部の松澤芳昭准教授（情報システム学），そのゼミ生，すでに原著を購入していた学部2年生の有志2名らとともに輪読会を行い，彼らから本書の内容や翻訳に関して貴重なフィードバックをいただくことができた。重ねて感謝申し上げたい。

また，原著者のSalganikは，訳者が意味のとれないところや誤記の疑われる箇所について問い合わせたところ，やりとりが複数回に及んだにもかかわらず，すべて快く答えてくれた。

本訳書が，デジタル時代の社会調査に携わる多くの人に活用され，日本における計算社会科学推進の一端を担ってくれることを切に願っている。

2019年3月

訳者一同

ビット・バイ・ビット
——デジタル社会調査入門
Bit by Bit: Social Research in the Digital Age

2019年5月10日　初版第1刷発行

著　者　マシュー・J.サルガニック
訳　者　瀧川裕貴（たきかわひろき）
常松淳（つねまつじゅん）
阪本拓人（さかもとたくと）
大林真也（おおばやししんや）
発行者　江草貞治
発行所　株式会社 有斐閣
郵便番号 101-0051
東京都千代田区神田神保町 2-17
電話　(03) 3264-1315〔編集〕
(03) 3265-6811〔営業〕
http://www.yuhikaku.co.jp/

印刷・大日本法令印刷株式会社／製本・大口製本印刷株式会社
©2019, H. Takikawa, J. Tsunematsu, T. Sakamoto, S. Obayashi. Printed in Japan
落丁・乱丁本はお取替えいたします。
★定価はカバーに表示してあります。

ISBN 978-4-641-17448-1

JCOPY 本書の無断複写（コピー）は，著作権法上での例外を除き，禁じられています。複写される場合は，そのつど事前に（一社）出版者著作権管理機構（電話03-5244-5088，FAX03-5244-5089，e-mail：info@jcopy.or.jp）の許諾を得てください。