CHAPTER 1

# Social Media, Sociality, and Survey Research

**Joe Murphy, Craig A. Hill, and Elizabeth Dean,**
*RTI International*

COPYRIGHTED MATERIAL

As survey researchers, we have long been concerned about the future of survey research. Beset by problems not of its own making, the survey research discipline faces unprecedented challenges because of declining data quality—stemming from, for example, falling response rates, inadequate sampling frames, and antiquated approaches and tools. Conducting a survey is, at its core, a social interaction between a researcher and a (potential) respondent—a "conversation with a purpose." The current pace of technological change—and the way people communicate with one another—presages the upheaval of survey research as we know it.

Thus, survey researchers should be—must be—searching for improvements to the way their research is conducted. Survey researchers should be—must be—constantly scanning the landscape of technological and social change, looking for new methods and tools to employ. In this spirit, we have been, somewhat jealously, watching the explosion of social media.

In this book, we introduce the concept of the *sociality hierarchy*; that is, three "levels" of sociality inherent in the current (and future) state of

*Social Media, Sociality, and Survey Research*, First Edition.
Edited by Craig A. Hill, Elizabeth Dean, and Joe Murphy.
© 2014 John Wiley & Sons, Inc. Published 2014 by John Wiley & Sons, Inc.

person-to-person interactions using computing devices: (1) broadcast, (2) conversational, and (3) community. Survey researchers should recognize these levels when attempting to apply new social media tools to survey research. This book presents examples of how this application can be done and, perhaps more importantly, how survey researchers should think about applying these tools in the future as a complement to traditional survey research.

In this first chapter, we discuss the advent of social media in its many and varied forms, and we define it from the perspective of a survey researcher. We also show why survey researchers should be interested in, and vigilant about, social media. We introduce the concept of the sociality hierarchy for social media and show examples of each level or category.

Throughout the rest of the book, we explain (and show), at a more practical level, how survey researchers can use the data generated by social media at each level of the sociality hierarchy. Finally, we suggest particular vectors on which survey researchers might find themselves with regard to the use of social media data and tools as we move inexorably into the future.

## WHAT IS SOCIAL MEDIA?

Millions of people have joined networks like Facebook and Twitter and have incorporated them into their daily lives, and at least partially as a result, communications between individuals and groups have changed in a fundamental way. Every day, billions of transactions occur over electronic systems, and within this stream are data on individuals' behaviors, attitudes, and opinions. Such data are of keen interest to those conducting survey research because they provide precisely the types of information we seek when conducting a survey.

The term *social media* has become ubiquitous. But what is social media? The term *social* suggests two-way interactions between people, which may be classified as one-to-one, one-to-many, or many-to-many. *Media*, or tools that store and deliver information, typically include materials that deliver text, images, or sound, i.e., *mass media* like books and magazines, television, film, radio, and *personal media* like mail and telephone.

Media can be further delineated into *analog media*, which contain data in a continuous signal or physical format, and *digital media*, which store information in a binary system of ones and zeros. The important distinction for our purposes is that digital media are typically stored or transmitted through computers or digital devices and can be disseminated via the Internet (aka the web).

The term *social media* most commonly refers to web-based technologies for communication and sharing over the Internet. There is no single agreed-upon definition of social media, but Scott and Jacka (2011, page 5) contend that social media "is the set of web-based broadcast technologies that enable the democratization of content, giving people the ability to emerge from consumers of content to publishers." Social media involves the intensive use of electronic media for people in contact through online communities (Toral et al., 2009), but no agreed-upon definition exists for the concept of "online community" either (De Souza & Preece, 2004).

We propose a specific working definition of social media for the purposes of survey research: *Social media* is the collection of websites and web-based systems that allow for mass interaction, conversation, and sharing among members of a network. In this definition, social media has four defining characteristics: user-generated content, community, rapid distribution, and open, two-way dialogue (Health Research Institute, 2012).

Social media must be distinguished from other similar terms that may refer more to the technological or structural aspects of online systems. For instance, the *web*, built on the infrastructure of the Internet, contains myriad sites that are not part of the social media construct. These sites include, for instance, information resources without an interactive, discussion, or sharing component. Such resources typify *web 1.0*, which describes resources that allow users to view and consume information from the web but without necessarily sharing or interacting with the contents (Krishnamurthy & Cormode, 2008). *Web 2.0*, however, refers to sites with user-generated content such as videos, music, blog text, and photos (Anderson, 2007). Web 2.0 has allowed users to interact with the web and with other users and has permitted the consumer to become the creator of content (Asberg, 2009). Web 2.0 has been used as an umbrella term for web-enabled applications built around user-generated or user-manipulated content (Pew Internet &

American Life Project, 2011). Ravenscroft (2009) considers web 2.0 to be the "social and participative web" that includes tools emphasizing social networking (e.g., Facebook, Bebo, and LinkedIn), media sharing (e.g., MySpace, YouTube, and Flickr) and virtual worlds (e.g., Second Life).

The distinction between social media and web 2.0 can be rather nebulous, but Figure 1.1 shows how social media fit into the evolution of popular media and technology. The figure presents analog, digital, web 1.0, and web 2.0 media types progressing over time from left to right. Analog media are listed with arrows connecting them to digital and web media that have evolved from or supplanted them. Diamonds represent media that have traditionally been employed for survey research including the following:

- Mail, used for delivering paper-and-pencil surveys
- Telephone, used for conducting survey interviews
- Digital audio, such as WAV files, used for recording interviews for verification and data quality purposes (Biemer et al., 2000)
- E-mail and homepages (now more commonly referred to as webpages) for web surveys, notifications, and panel maintenance
- Market research online communities, which are private social networks and websites to conduct qualitative marketing research with a selected community of members.

Media represented by hexagons in Figure 1.1 hold potential for survey research innovation moving forward and are, thus, the focus of this book. In this book, we present our research on these new modes and methods such as online games, multiply massive online games, and virtual worlds (Chapters 5, 6, 10, and 11); social networking platforms like Facebook and Twitter (Chapters 2, 3, 4, and 9); smartphone-enabled media like mobile apps for survey data collection (Chapter 7); and web 2.0-enabled research resources such as crowdsourcing (Chapter 8).

The dotted line in Figure 1.1 indicates which media we consider *social media* for the purposes of this book. It encompasses web 2.0 media but also includes some typically classified under web 1.0 that allow for user-generated content, rapid distribution, or open, two-way dialogue.

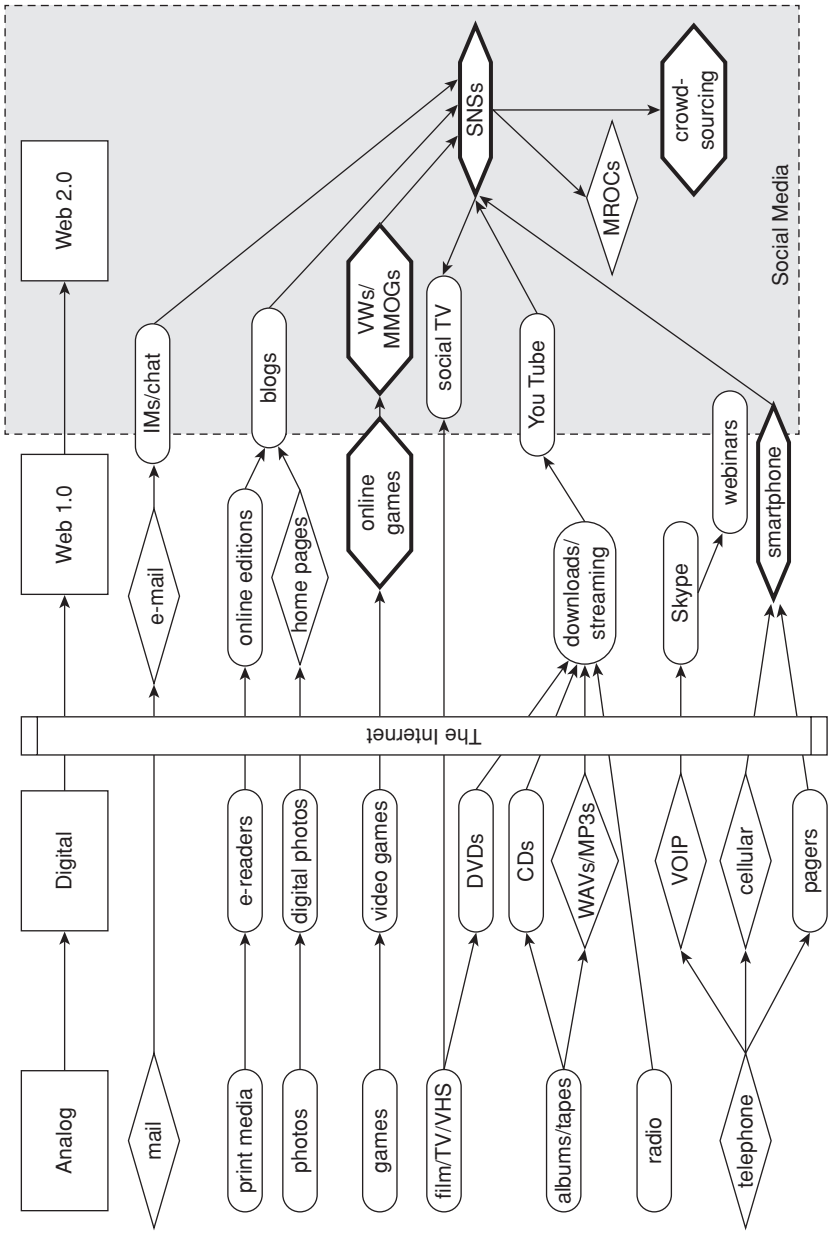


**FIGURE 1.1** Media types.

## SOCIAL MEDIA ORIGINS

Consonant with our definition of social media as web-based, most social media originated—were born—in the 1990s. The explosion in popularity of the Internet during this time eventually led to its steady penetration into many aspects of life (Leberknight et al., 2012). Between 1993 and 1995, Internet service providers (ISPs) began offering access in most major U.S. cities (Scott & Jacka, 2011). America Online (AOL) became a popular service by mailing access discs directly to consumers, which allowed them to try (and subscribe to) the Internet. Once on the Internet, users could participate in social activities such as sharing their opinions broadly through individual homepages, participating in bulletin board discussions, and engaging in other activities. Widespread use of social media and web 2.0, though, did not proliferate until the 2000s. With the launch of MySpace in 2003, users had the means to control and share media easily on their own personal pages and comment on the contents of their contacts' pages. Scott and Jacka (2011, page 14) argue that "if 2000–2004 was about building platforms and tools, 2005–2009 could be defined as the period of user adoption and the remarkable change in how users connect, converse, and build relationships." By 2006, YouTube and Twitter had launched, and Facebook cracked the 10-million user mark.

Since 2006, the function of the web has moved rapidly in the direction of user-driven technologies such as blogs, social networks, and video-sharing platforms (Smith, 2009). This user-generated content is becoming more prevalent across the web, with most sites now allowing users to publish opinions, share content, and connect with other users.

## SOCIAL NETWORKING SITES AND PLATFORMS

Social media use has skyrocketed in the last several years. Participation in social networking sites ballooned from 5% of all adults in 2005 to 50% in 2011. Facebook, in particular, grew from 5 million users in 2005 to more than 900 million in 2012. Twitter continues its amazing growth with nearly 500,000 new accounts created per day (Health Research Institute, 2012).

Web 2.0 tools and technologies in particular have allowed collaboration and communication across boundaries (Schutte, 2009). At the core of web 2.0 are social media sites like Facebook, Twitter,

LinkedIn, YouTube, and many others facilitating social sharing to an unprecedented extent. These sites can be defined as *social network sites* or *social network services*—"web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system" (boyd & Ellison, 2007, page 211). On social network sites, individual users share knowledge, pose and solve problems, seek and offer advice, tell stories, and debate issues of interest (Toral et al., 2009). boyd and Ellison differentiate social network sites from *networking* as the latter term implies that relationships are being initiated between strangers. Social networking sites make networking possible, but it is not their primary function nor is it what makes them distinct from other forms of computer-mediated communication. The defining quality of social network sites, according to boyd and Ellison, page 211, is that they "enable users to articulate and make visible their social networks."

Social networking sites host online communities of people who share interests and provide ways for users to interact, including e-mail and instant messaging services (Shin, 2010). Social network sites allow users to create profile pages with personal information, establish "friends" or contacts, and communicate with other users (boyd & Ellison, 2007). On popular sites like Facebook, users communicate via private messages or public comments posted on profile pages. Users also have the option to communicate via instant message, voice, or video. Early adopters of the sites tended to be young people, but an increasing proportion is older. And, indeed, extroverts may use social network sites more often (Sheldon, 2008) as these sites contain relatively public conversations between friends and contacts (Thelwall et al., 2010).

In the United States, 81% of adults go online and 67% of online adults visit social networking sites. Forty percent of those who own a cell phone use a social networking site on their phones (Pew Internet & American Life Project, 2013). Among teens aged 12 to 17, 93% are online and 74% have a profile on a social networking site; social networking is the online activity on which they spend the most time daily (Pew Internet & American Life Project, 2011; Kaiser Family Foundation, 2010). Popular social networking sites and platforms take many different forms, as described in the next section.

## Blogs

Blogs (or web logs) are periodically updated sites managed by an individual or group to provide information or opinion on a range of topics. Entries, or posts, are organized so that the most recent entry appears at the top. Many blogs allow and invite discussion from readers in the form of comments directly on the site. In popular blogs, the discussion via comments can contain more varied information and opinion than the original blog post itself. Blogs vary in length, frequency of post, topic, and formality. Because of the advent of free and easy-to-use web publishing tools (e.g., Blogger, WordPress, and Tumblr), blogs are available and used as an outlet for those with a wide range of resources and proficiency with computing. More than 150 million public blogs exist on the web (NielsenWire, 2012). Compared with the general population, bloggers are more likely to be female, in the 18- to 34-year-old age group, well-educated, and active across social media.

Blogs are often interconnected, and writers may read and link to other blogs, providing links to others in their entries. Often referred to as the blogosphere, interconnected blogs are social media in their own right, with social communities and cultures of their own (Mazur, 2010).

## Twitter

Twitter has been defined as a "microblogging" service. Launched in 2006, Twitter users or "Tweeters" publish short messages, or "Tweets," up to 140 characters in length. Tweets are publicly visible through the website or third-party applications in real time with millions of Tweets posted across the world per hour. Tweets contain status update information, sharing of links and information, direct messages to other Tweeters, and opinions on almost any topic imaginable (Tumasjan et al., 2010).

With the rise of text-based social media like Twitter and millions of people broadcasting thoughts and opinions on a variety of topics, researchers have taken notice. Twitter, in particular, is convenient for research because of the volume of publicly available messages and the process of obtaining them is relatively simple (O'Connor et al., 2010).

Survey research applications for Twitter are described in Chapters 2, 3, and 9.

## Facebook

Launched in 2004, Facebook is a website and social networking service with almost 1 billion active users as of 2012, making it one of the most popular sites on the Internet. Enrollment is free, and users can create a profile and share information such as hometown, current city, education, employment, interests, and favorites. Users can also post photos, videos, notes, and status updates to share with their Facebook "friends." Through Facebook, a user can make new contacts and follow, or "like" different groups, organizations, or products. A survey of 1,487 adults in the United States found that 43% of U.S. adults use Facebook; use is highest among young adults, college graduates, and those with an income greater than $90,000 (Morales, 2011).

Facebook users and their "friends" comprise virtual communities linked by shared interests or opinions. Researchers can sample populations of interest working through existing groups or creating new ones for the purposes of their analysis (Bhutta, 2012). As the number of Facebook users worldwide continues to grow, survey researchers are interested in how we might leverage social media to connect individuals and share information with one another (Lai & Skvoretz, 2011). Potential survey applications using Facebook are described in more detail in Chapter 4.

## LinkedIn

Often considered the business world's version of Facebook, LinkedIn is a social networking service for professionals. With more than 150 million users in over 200 countries and territories, LinkedIn provides a means for individuals in various professional fields to interact and discuss issues, build their personal networks, and promote their expertise when seeking new opportunities or employment. Groups serve as hubs of interaction with sharing of announcements, links, and opinions. Over 47 million unique users visit LinkedIn monthly with more than 21 million from the United States.

## Second Life

Second Life is an online three-dimensional world in which users (called "residents" in Second Life) design personal avatars and interact with

other avatars and their surrounding environment. Communication in Second Life can be through instant messages or voice chat. As opposed to other social media sites like Facebook and Twitter, which typically augment real-life personas and relationships, residents in Second Life represent themselves in ways that depart from real-life appearances and personalities.

Second Life residents come from more than 100 countries. Second Life use is measured in user-hours—and 481 million user-hours were logged in 2009 (Linden, 2011). The most active users (as of 2008) were 25–44 years old (64% of hours logged) and male (59% of hours logged). Second Life provides a context-rich environment for conducting cognitive interviews and other survey pretesting activities (Dean et al., 2009; Murphy et al., 2010). The system allows the researcher to target and recruit specific types of residents through classified-type advertisements, online bulletin boards, and word-of-mouth in the virtual world, which can be more efficient and cost-effective when compared with traditional newspaper ads or flyers that are typically used to recruit in-person cognitive interview subjects (Dean et al., 2011). Text-based chat and voice chat can be collected for analysis, offering full transcripts of cognitive interviews (see Chapter 5). The only elements of in-person cognitive interviews missing from the paradata are facial and physical expressions, although Second Life residents can manipulate these for their avatars to a certain extent. Second Life is also useful in accessing hard-to-reach populations such as those with chronic illnesses (see Chapter 10) and for conducting methodological research (see Chapter 6).

## Other Social Networking Platforms and Functionalities

A multitude of other social media platforms and functionalities exist, and the list grows continually. YouTube (owned by Google), for example, is a website for sharing videos. Users can upload original content (e.g., video blogs and performances) or media clips and discuss content on the site using comments. Flickr, Snapchat, and Instagram are popular sites with a similar design but primarily for posting and sharing photographs.

Another social networking site, Foursquare, focuses on physical locations. Users "check in" with their global positioning system

(GPS)-enabled smartphones at various locations and earn awards in the form of points and badges.

For the purposes of this volume, we consider text messaging and smartphone applications (i.e., mobile apps) (although not technically web-based) to fall under the umbrella of social media. Text messaging can be used to interact with study respondents to capitalize on the cultural behaviors of the younger population. A text message is a short communication between mobile phones on a bandwidth lower than that of a phone call, and messages are usually limited to 160 characters (Cole-Lewis & Kershaw, 2010). Mobile apps are standalone pieces of software that a user can download to his or her device that allow for communication and interaction in many ways similar to web-based social media platforms. Chapter 7 discusses mobile apps in more detail.

## WHY SHOULD SURVEY RESEARCHERS BE INTERESTED IN SOCIAL MEDIA?

### The Current State of Survey Research

Survey research, as a discipline, has been under attack for the better part of several years now because of several coincidental trends that imperil the ability of a survey to provide data that are fully "fit for use"—accurate, timely, and accessible (Biemer & Lyberg, 2003, p. 13). One could argue that, in this day and age of advancing technology, survey data can be made *timelier* and *more* accessible quite easily—but if those data are not accurate, they are not of high quality and are not "fit for use."

These coincidental developments and trends have resulted in an uneasy state of affairs for survey researchers. We now confront declining data quality as a result of, among other issues, falling response rates and inadequate sampling frames.

### Falling Response Rates

Virtually everyone in the survey research industry knows that, regardless of mode, survey response rates (the number of completed interviews divided by the eligible sample population) are in decline. Scholars in the survey research field have been decrying the decline since at least the 1990s (de Heer, 1999; Steeh et al., 2001; Tortora, 2004; Curtin et al., 2005). That outcry continues to this day: In his presidential address

delivered to members of the American Association for Public Opinion Research (AAPOR) in May 2012, Scott Keeter called the problem of declining response rates "formidable" and noted that they are "well below the norms we were trained to expect." Furthermore, he cited response rates in the single digits in the public opinion (polling) sector of the industry even when conservatively calculated (Keeter, 2012, page 601).

Why are falling response rates a problem or threat to traditional survey research? Data quality has long been a cornerstone of social science research: Researchers are constantly investigating methods to maximize data quality and minimize survey error, which has many components and causes. The failure to obtain data from all sample members, referred to as unit nonresponse, can lead to bias in a researcher's estimates and, thus, flawed conclusions resulting from analysis of survey data if systematic differences exist in the key survey outcomes between responders and nonresponders (Peytchev, 2012). If, for example, the underlying cause for increased unit nonresponse is the *topic* of the survey (e.g., many sampled respondents refuse to participate in a health survey because they view their own health as a sensitive topic), the level of unit nonresponse (the response rate) will almost certainly be correlated with key survey estimates (Groves et al., 2006).

Although researchers and analysts can assess nonresponse bias and its impact in multiple ways, the response rate is the most well-known and most frequently used indicator of error stemming from unit nonresponse. It has been, and remains, "an overall indicator of data quality—interpreted by many as a shorthand way for labeling whether a particular survey is "good," "scientific," "valid"—or not." (Carley-Baxter et al., 2006, page 2).

A combination of factors has likely led to survey researchers' increasing inability to achieve adequate response rates, including increased suspicion regarding requests for data from both government and corporations, increased reluctance to share personal data with unknown persons or entities, and tightened controlled-access to housing units, among others.

Some of the mistrust by potential respondents is no doubt because of burgeoning bids for their attention (and answers/data) from every conceivable angle: The amount of junk mail (both print and electronic) is rising. Telemarketers discovered, early on, the power and ease of calling people at home to ply their wares—so much so, that Americans

were inundated with calls that were indiscernible at first from survey requests (Tourangeau, 2004). Some people might have sought to escape the onslaught by declining to answer the door or phone, and instead seeking refuge in their computer, but then spam e-mail and phishing attacks became more prevalent than ever (Kim et al., 2010; Tynan, 2002). The threat of computer viruses from unknown sources, and news stories about identity theft and stolen laptops containing confidential information likely led people to become more protective of their personal information, including refusing to cooperate with a survey solicitation—not bothering to make the distinction between a legitimate survey and any other entreaty for personal information.

People now have all sorts of ways by which they can avoid unwanted solicitations, including requests to participate in surveys. Mobile phones come equipped with caller ID. If people do not recognize the number, they can easily reject the incoming call. Such technology has certainly had an impact on surveys, contributing to declines in response rates and increasing the costs of conducting telephone surveys (Kempf & Remington, 2007; O'Connell, 2010).

Challenges with nonresponse have not been limited to telephone surveys. In-person (or face-to-face or field) surveys have been the gold standard for survey research since at least the 1940s, and most of the massive data collection efforts required by the U.S. federal statistical system are conducted this way (Yeager et al., 2011). But even gold standard surveys are experiencing downward-trending response rates. One reason is the increase of controlled-access housing developments, such as gated communities and buzzer systems in multiunit housing, making it much more difficult for even the most persistent interviewer to contact potential respondents (Keesling, 2008). In recent years, government-funded studies have documented an increase in the amount of nonresponse attributed to controlled access (Cunningham et al., 2005; Best & Radcliff, 2005). Because individuals have been able to restrict access in these ways, reaching them and collecting quality survey data has become more difficult and more costly (Curtin et al., 2005).

## Frame Coverage Errors

A second major potential source of error in survey research—and, thus, another nail in the coffin being built by its detractors—is the ever-increasing number of cell-phone-only households and individuals.

Surveys conducted by landline telephone, using a random-digit-dial (RDD) sampling frame, run the risk of missing entire segments of the population of interest if the traditional landline-based telephone sampling methods are not combined with a cell-phone frame (Blumberg & Luke, 2009). The latest data from the U.S. federal government show that more than a third (36%) of American adults now has a cell phone but no landline (see http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201212.pdf or Blumberg et al., 2012). Some subgroups are more likely to fall in the cell-phone-only category: adults aged 25–29 (58.1%) and renters (52.5%), for example. These holes in sampling frame coverage have led some researchers to suggest that we abandon telephone sampling and telephone surveys entirely (Holbrook et al., 2003) or that the "entire industry needs to be remade" (O'Connell, 2010).

Other scholars are trying to develop alternative schemes for sampling, acknowledging that something new must be done (see Brick et al., 2011; Link et al., 2008). In this book, we suggest that many social science researchers are now moving beyond alternative sampling schemes and are instead exploring entirely new models of social science data acquisition and analysis. At least some of this search for new ways to find and interpret data is driven by the ready accessibility of data itself: The era of "Big Data" has begun.

## The Coming Age of Ubiquity

In Alvin Toffler's *The Third Wave* (1980), he writes that there have been three waves of civilization, beginning with the agricultural wave, followed by the industrialization wave, and now supplanted by the information wave. Toffler further argues that the advent of each wave brings with it uncertainty, upheaval, and change. Those willing to surf along on top or ahead of the new wave—that is, embrace the change—can not only acquire power and wealth, but also fashion fundamental alterations of civilization.

The ubiquity of computers and technology has been the stuff of science fiction for decades. This ubiquity seems tantalizingly close now, but they are still just out of reach. The day is coming when your refrigerator "knows" that it is out of milk and signals the local grocery to add a gallon of 2% milk to your next order. We will not, however, have achieved this state of ubiquity until we no longer take notice of the role that computers and technology play in our everyday lives; that is, implicit in the notion

of "ubiquity" is the idea that computers will perform these functions as a matter of course and we will have accepted that.

*Variety.* Transactions and interactions that are heavily computer dependent already occur on a daily basis. Making a phone call, sending a text message (SMS), making a bank deposit, swiping a loyalty card at Starbucks or Walmart, paying for a meal with a credit card, passing in front of a CCTV camera in London, posting a video on YouTube, querying Google, Tweeting on Twitter, and a hundred other seemingly innocuous activities that we all perform every day all involve a vast network of computers humming away behind the scenes, generating vast amounts of data. All of these bytes of information posted or transacted each day, taken together, equal "Big Data."

*Volume.* As computers and computer-based technology near ubiquity, the amount of transactional and interactional data created grows exponentially. "In 2012, every day 2.5 quintillion bytes of data (1 followed by 18 zeros) are created, with 90% of the world's data created in the last two years alone," according to one estimate (Conner, 2012, paragraph 3). In February 2011, Facebook users uploaded, on average, 83,000,000 photos *every day*. Walmart generates 1 million customer transactions every *hour*. Every *minute* (http://mashable.com/2012/06/22/data-created-every-minute/), for example, there are:

- 47,000 app downloads from the Apple App Store
- 3,600 pictures uploaded to Instagram
- 27,778 new posts on Tumblr
- 204,166,667 e-mails sent

*Velocity.* Computers operate at an ever-increasing pace, contributing to the generation of more and more data. Any particular set of variables of interest may now be analyzed on a daily basis, or even hourly or by the second, if preferred. Faster data processing and faster data transmission rates make for shorter and shorter feedback loops, making data and data analysis timelier than ever before.

*Validity.* However, we know that social media has produced much superfluous material as well. People create Facebook pages for their cats, sensors capture false positives, and data elements go missing. One of the biggest roles that social science researchers can play in this arena is to attempt to make sense of it all: Revolutions in science have often

been preceded by revolutions in measurement, and the explosion of data and data accessibility demands new ways of measuring human behavior (Cukier, 2010).

There is a distinction to be made between the "organic" (Groves, 2011) Big Data that are created at a volume of billions of gigabytes per day and the "designed" data that survey researchers have been collecting for nearly a century: Survey methodology, as a field, has focused on developing less costly and more accurate methods of asking people questions so as to obtain self-reports of behaviors and opinions. Those answers populate designed datasets, produced to answer a very specific set of research questions. Those "designed" data are case-poor and are observation-rich in comparison with Big Data; in other words, to control costs, we sample the population of interest and ask a specific set of questions, which results in a dataset with (relatively) few cases, but with just the desired data elements for each case. Organic data, on the other hand, emerging out of new communication technologies and without any specific research purpose or design, grow organically and without boundaries. Separate pieces of data, collected for different reasons and by different entities, can be combined to make even bigger datasets that have the capability to answer questions that neither original dataset could. The result is billions of "cases," but for researchers, the data elements are ill-defined. Nonetheless, this boundless—and boundaryless—set of data has too much potential to ignore: The temptation to use these data to understand social behavior and trends is nearly irresistible. The promise and potential of Big Data, of course, are that they can answer Big Questions. We are on the cusp of an era, a wave, where data are cheap and abundant instead of, as they are now, scarce and expensive to obtain; in fact, a considerable amount of Big Data is actually free to obtain (or, at the very least, so inexpensive as to be an insignificant cost). As the amount of cheap or free Big Data increases, its accessibility (one of the components of data quality, as described above) also goes up, driving its overall value ever higher. One fundamental change that results from these easily accessible and plentiful data is that a business model based on "selling" data will not be sustainable; instead, companies will have to adjust their business model so that what they are selling is analyzed and interpreted data—not the raw data alone.

Perhaps another fundamental change is that, to add value to a set of data, many social science researchers will have to change the way they approach a dataset. Traditionally, survey researchers, statisticians, and data analysts have been accustomed to working with a (relatively

speaking) small dataset to which a complex model is applied. With Big Data, we will see more unstructured models applied to huge datasets.

Instead of asking questions of a sampled population, as survey researchers have done for decades, we may now be asking questions of data—instead of people—that already exist. The goal for survey researchers, statisticians, and data scientists as they face the challenge of understanding, analyzing, and drawing conclusions from Big Data will be the same as it has always been for survey researchers: "producing resources with efficient information-to-data ratio" (Groves, 2011, page 869).

Still, at least for the immediate future, interpreting the results of these unstructured models will involve art as well as science. Already, the entire fitness-for-use dynamic is changing: Data, now coming from (or in combination with) "Big," organic sources, instead of just survey (or designed) data, are timelier and more accessible than ever. Social science researchers and survey researchers have a unique opportunity to define and develop the "accuracy" component of the fitness-for-use definition of quality.

## Public vs. Private Data

Most Big Data are relatively inexpensive, but there are still questions to be answered about how accessible these data will be. Do we, as researchers, have unfettered and unlimited access to all these Big Data? Although we think that such a complicated issue does not lend itself to an easy answer, others think there *is* an easy answer: "You have zero privacy anyway," said Scott McNealy, the CEO of Sun Microsystems. "Get over it" (*Wired*, 1999, paragraph 2).

Facebook, for example, has come under fire for repeatedly changing its privacy policies (Terms of Use); as of this writing, the policy has 115 clauses, 25 subclauses, 10 appendices, and 13 links to other documents/policies (Facebook, 2012). Most Facebook users likely do not fully and completely understand who can see their latest status update or check-in.

Many companies sell products or services now with the full expectation that the data generated by the product will belong to the company. Nike+ FuelBand and Fitbit, for example, provide users with data about calories burned during exercise, helping users to track their activity level and weight. The device also provides that data to the manufacturer. Other companies (e.g., MC10 and Proteus) make products that monitor heart

rate, body temperature, and hydration level. Companies see the value in these (anonymized) data that, once captured, will be quite interesting to healthcare providers (Hardy, 2013).

Social networking sites make public many personal pieces of information about users. Profiles may contain users' name, age, gender, geographic location, and information about hobbies, interests, attitudes, and opinions (Wilkinson & Thelwall, 2011). Most sites have privacy settings that allow users to control who can access their information, but many users leave their profiles public, either by choice or because they do not take the time to understand the distinction, which allows anyone to access this information. As Wilkinson and Thelwall (2011, page 388) explain, "In theory, this gives academics and market researchers unparalleled access to mass personal data with which topics like homophily (Thelwall, 2009; Yuan & Gay, 2006), depression (Goh & Huang, 2009), brand perception (Jansen et al., 2009), and personal taste (Liu, 2007; Liu et al., 2006) can be investigated relatively cheaply and easily."

Social networking sites and platforms allow mass-scale sharing of thoughts, opinions, and behaviors by people worldwide. As a result, a potential goldmine of data is easily accessible, and these data could be harvested to provide insights in the manner that survey data have traditionally been employed.

## SOCIAL MEDIA INTERACTION: NEXT WAVE (OR SUBWAVE)?

All of the aforementioned developments suggest that we are approaching the next phase of evolution of survey research. Survey research modes emerge over time as communication technologies and preferences change. Up to this point, the dominant survey research modes have been face-to-face, mail, telephone, and web. Each point along this evolutionary pathway has spawned a new body of research and new methods for increasing the reliability and validity of survey data, thereby reducing the error in the data collected, and increasing its fitness-for-use (quality). As survey researchers with a keen interest in how communication technology shapes the way we ask questions and record answers, we want to know the answers to these questions: Are social media and mobile technology the next wave in survey research? Are these changes a subwave of Toffler's Third Wave? And if they are the next wave or subwave, is the survey research community ready for it?

The history, or evolution, of survey research shows several times at which the research community decided collectively that new modes were acceptable. Early on, survey sampling proponents took advantage of the convergence of (1) well-defined area probability sampling ingredients like census tract information, (2) trends in contemporary print media favoring sharing of individuals' opinions as part of news stories, and (3) an emergent cognitive science that provided tools (questionnaires and scales) for measuring individuals' behaviors and attitudes. These three factors were combined with the communication tools readily available at the time (mail and in-person interaction) to develop the survey interview protocol as we know it (Groves, 2011).

But, the 1960s introduced a departure from that evolutionary vector. Overall, by 1958, 73% of the U.S. population had telephones (Mendelson, 1960), so survey researchers began to consider the telephone as a possible new mode. In 1960, *Public Opinion Quarterly* published a summary of a roundtable discussion from the 15th meeting of the American Association of Public Opinion Research (AAPOR), *Current Status of Telephone Surveys*. In this discussion, presenters pointed out concerns with telephone coverage among less affluent populations, nonwhite households, and rural areas. Then, in a 1964 study of white women of child-bearing age in the Detroit area, 91.3% of respondents were able to be interviewed by telephone rather than by face-to-face or mail methods (Coombs & Freedman, 1964), helping to convince the research community that the telephone was a viable, valid mode. Coverage grew from 81% in 1963 to nearly 95% in 1983 (Thornberry & Massey, 1988). By eliminating the time and costs of travel for face-to-face surveys and reducing the lag time of mailed surveys, telephone as a mode of data collection increased the efficiency with which survey data could be collected, making survey data both more accessible and more timely.

In the 1990s, another possible departure from the evolutionary pathway took shape. The Pew Research Center for People and the Press first began tracking Internet adoption in June 1995, at which time 14% of adults had access (Pew Internet & American Life Project, 2012). In 2000, Don Dillman (2000) published *Mail and Internet Surveys: The Tailored Design Method* and Mick Couper (2000) published the *Public Opinion Quarterly* piece, "Web Surveys: A Review of Issues and Approaches," which together signaled the beginning of the era of web survey design. By February 2001, Pew reported that 53% of American adults had access to the Internet, and now, Internet penetration in the United States is currently at 81% of adults (Pew Internet & American

Life Project, 2012). The Internet/web is now used readily as a data collection mode, but coverage problems (the "digital divide"), like those of the telephone in the 1950s and 1960s, have prevented it from completely replacing current modes. Instead, Internet data collection is now part of the multimode arsenal or approach adopted by most survey research practitioners to combat the response rate/nonresponse bias problem.

Nevertheless, Internet penetration—and usage among the connected continues to grow. Our definition of social media encompasses a broad range of communication technologies: Social networking sites predominate, but virtual worlds, mobile technologies, blogs and microblogs, Internet telephony services like Skype, and crowdsourcing platforms all fall under the umbrella of our definition. As shown in Figure 1.2, 69% of adult Internet users spend time on social networking sites like Facebook, LinkedIn, or Google Plus; 71% watch videos on sites like YouTube or Vimeo; 53% search online classified ads sites like Craigslist; 46% upload photos; 25% use an online calling service like Skype; 16% use Twitter; 14% author blogs; 8% use online dating services; and 4% use a virtual world like Second Life.

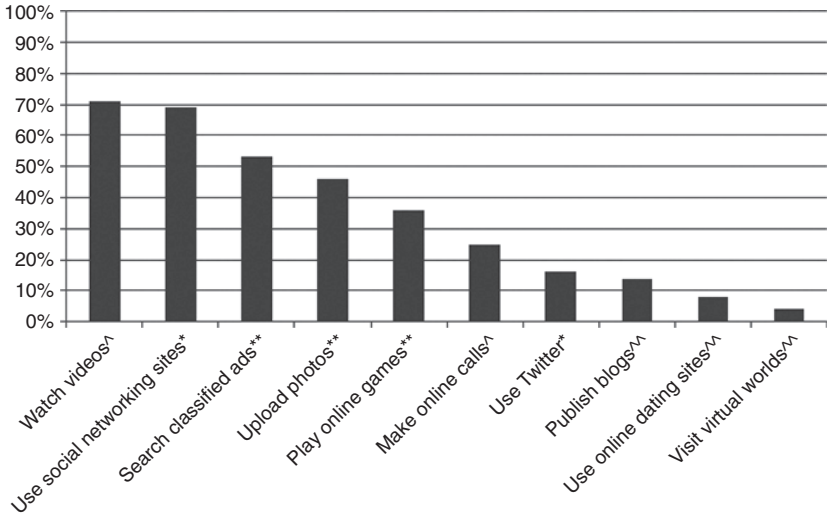


**FIGURE 1.2** Percentage of online U.S. population engaging in Internet activities. ^Data last collected in 2011. *Data last collected in 2012. **Data last collected in 2010. ^^Data last collected in 2009. *Source:* Pew: http://pewinternet.org/Trend-Data-(Adults)/Online-Activites-Total.aspx.

Are these rates of adoption and use sufficiently high to be considered the next (sub)wave? Do social media now represent a new mode for survey research? All of these activities generate Big Data, which as noted, results in timely and accessible data. It is time to consider adding methods of capturing and analyzing these data to the survey researcher's toolbox.

## ADDING SOCIAL MEDIA TO THE SURVEY RESEARCH TOOLBOX

Already, a plethora of approaches exist that use social media as an adjunct to, or even a replacement for, survey research. Not all of these approaches possess methodological rigor, and certainly, not all find favor with full-probability sampling adherents. Suffice it to say, many techniques are now available for harnessing social media content or social media communication methods as a way of adding to our understanding of human social behavior. Blog posts, Facebook status updates, YouTube comments, online dating or professional profiles, and Tweets all provide passive content (Big Data) about individuals that can be used to supplement our understanding of public opinion. These types of data are passive because researchers do not actively collect them; rather, the data are extant, provided by users, and are available to be analyzed.

For example, researchers can use multiple methods to monitor and mine social media content. Chapters 2 and 3 include two excellent examples of analysis of publicly available social media data, using sentiment analysis, a subfield undergoing rapid growth. Sentiment analysis assigns positive, negative, or neutral codes to text, enabling conclusions about the public sentiment on a particular topic or issue. Researchers can also apply content analysis to passively collected data, going a step beyond sentiment ratings and, instead, looking to uncover recurrent themes in the data. For example, a content analysis of a Twitter conversation about concussions found that 37% of concussion-related content was about recent news stories on concussions, 27% involved a user sharing a personal story or experience relating to concussions, and 13% of content included recommendations for or descriptions of treatment procedures (Sullivan et al., 2011).

Another method of online sentiment and content analysis is called netnography (Kozinets, 2010): ethnographic research that takes place in

online communities. Netnographers may use content or sentiment analysis to understand the communities, but as an ethnographic method, it involves some interaction with research subjects; the researcher becomes a participant in the online community to better understand its culture and content. Both content and sentiment analysis may be used to analyze the results of netnographic data collection.

Netnography bridges the gap between passive data collection and the investigator-initiated, more traditional way of eliciting data from participants. And myriad other opportunities exist for data collection from social media platforms beyond immersive cultural studies. In the chapters that follow, survey samples are developed through a mobile panel recruited from iTunes and Google Play, via two crowdsourcing platforms, and from online ads placed on Craigslist and Facebook, from using hashtags on Twitter, and within Second Life user forums. Questions are asked in 140-character Tweets, by avatars inside an immersive three-dimensional (3-D) virtual world, within a Facebook app, within a mobile survey app, as tasks assigned through a crowdsourced company, and as part of games designed to make the process of answering questions more fun.

All of these approaches can be legitimate methods of producing quality data that are timely, accessible, and accurate—although, to be sure, many issues are still to be debated with these new and evolving methods. One way of sorting through the choice of which method to use is to think carefully about the nature of the data to be analyzed. Were these data arrived at passively, or were they produced by direct questioning? Were these data broadcast to a large and impersonal audience, or were they part of one-to-one communication? Did these data come from an online community of like-minded individuals who, collectively, represent some sort of group? Sociality can be a useful concept for cataloguing these data.

## TOWARD USING THE CONCEPT OF SOCIALITY IN SURVEY RESEARCH OF THE FUTURE

Sociality can be defined as the extent to which an individual is social or sociable, or the tendency to associate with groups (*The Free Dictionary*). Put another way, the extent to which a person socializes, or interacts with others, is *sociality*. People are inherently social, and they associate

with others and form groups for the simple goal of forming social relationships (Fiske, 1992). Thus, human sociality involves establishing identity, organizing social behavior, and convening and communing with other people; it also includes various levels of relationships, including private, public, and family (LaMendola, 2010). Survey researchers recognize that data collection takes place in a social context and, thus, takes into account sociality, whether expressly acknowledged or not. To be sure, sociality can be observed in many stages of the survey interview. Sociality can affect surveys in the following ways:

- Different experiences among respondents' family and relationship dynamics might result in different interpretation of questions about household income.
- Field interviewers observe neighborhood socioeconomic status, coming-and-going times, and friendliness/openness when deciding the best way to approach households selected for data collection.
- Survey designers take steps to establish privacy in the interview setting (self-administration, no identifying names, and private rooms) before posing sensitive questions to respondents.
- Asking questions about sexual behavior in the context of a survey about drug abuse and other risky behaviors might result in different answers than when asked in a survey about family and marital relationships.

The growth of online channels of communication and the resultant social media interaction presage an expansion or alteration of the social context, or sociality, of survey research. If human sociality provides the context in which survey research takes place, then the growth of online sociality provides an entirely new context with new ramifications for survey research.

For example, many research questions in surveys concern private behavior. The next subwave (of the Information Wave) has the potential to alter radically the sociality of a survey research interview. If people are now willing to disclose more private information, as they do on social media sites, they may be more willing to disclose it to researchers, especially if we adopt (or co-opt) new tools for eliciting that information. As multiple examples in this book demonstrate, people are quite willing,

using various social media channels, to broadcast information, thoughts, motivations, and ideas that were previously considered private.

Human sociality is both restricted and enhanced by online communication technologies. On the one hand, these new technologies or communication platforms facilitate the construction of identity (profiles), manifest relationships (friend networks/lists), and enable community (within apps, Facebook groups, blogs, and message boards). On the other hand, online social spaces present quite real challenges to individual privacy and boundaries. In an environment so open and oriented toward sharing, it becomes difficult to manage an appropriate level of disclosure for all different types of human relationships (personal, professional, public, family, romantic, and friendly) (Leenes, 2010).

Social network services (SNSs) like Facebook specifically encourage informal forms of sociality including chat, expression of tastes and preferences ("likes"), and gossip. As communication tools, SNSs turn previously informal or chat-type discourse into a formalized record with an audit trail; ultimately, certainly, such trails will change the norms of social behavior/sociality. Data points previously kept private, such as relationship status or sexual orientation, are now aired in public via SNSs and other online profiles. Although SNS users can choose what to disclose, and they can customize privacy settings, any post, entry, or status update has the potential to be disclosed when an SNS's privacy policy is changed (van Dijck, 2012). Despite these privacy concerns, however, many individuals clearly continue to broadcast their innermost thoughts (no matter how profound or mundane) using SNS platforms.

At another level, the expansiveness of online sociality has the potential to change how survey researchers engage respondents in a "conversation with a purpose." The survey interview has long been defined as such (Bingham & Moore, 1924; Cannell & Kahn, 1968; Schaeffer, 1991), but now, perhaps, the limitations (and advantages) of conversing by a tweet or with an avatar in a virtual world suggest that we may be able to redefine the survey research conversation.

As a final example, the growth in online communities—shared conversational environments and interfaces—portends not only another way for sociality to be expressed but also another way to conduct survey research conversations.

This hierarchical way of conceptualizing individuals in their social contexts is not unique to our application of sociality. Our model maps to the social–ecological model (SEM), a tool developed by the Centers

for Disease Control and Prevention (CDC) to develop health intervention and prevention programs to address health concerns at multiple levels (McLeroy et al., 1988). The SEM has four levels: individual, relationship, community, and societal, which are interrelated (Bronfenbrenner, 1994). The individual level includes personal and biological traits that contribute to a particular disease, condition, or behavior. The relationship level includes characteristics of family, peer, and romantic relationships that affect health status. The community level includes the setting and environment in which these relationships and individual behaviors occur (for public health considerations, these environs could include neighborhoods, towns, schools, and work environments, among others). The societal level includes social forces that affect health status, including social norms, public policies, and economic markets (Elastic Thought, 2008). The SEM model has been applied to public health problems as diverse as bullying (Swearer, Napolitano & Espelage, 2011), violence (Dahlberg & Krug, 2002), and child abuse (Belsky, 1980).

We now use this hierarchical way of conceptualizing individuals in their social contexts to help understand the impact of online sociality on communications and survey research. Figure 1.3 depicts the sociality hierarchy, a model that has three levels of online communication: broadcast, conversational, and community-based conversation. Each level has different implications and opportunities for survey research via social media and other new communication technologies.

*Broadcast social media* occur at the individual level. At this level, online users speak to the crowd or from a virtual soapbox, dispensing information about themselves for anyone to consume. Perhaps the best examples of broadcast behaviors are "Tweeting," sharing information via public profiles and blogging. The essence of *broadcast* social media, as we define it, is one person communicating with many others.

*Conversational social media* occur at the interpersonal level, that is, between two people, similar to analog or face-to-face communication. At this level, the dominant form of conversation is dialogue or one-to-one communication. Examples of conversational behaviors using new technologies and social media include videoconference calls via Skype, sending a text message to a friend, or having a conversation with another avatar in Second Life.

*Community-based conversation* via social media occurs when groups communicate with each other and within their membership ranks. Examples include Facebook groups organized around a professional

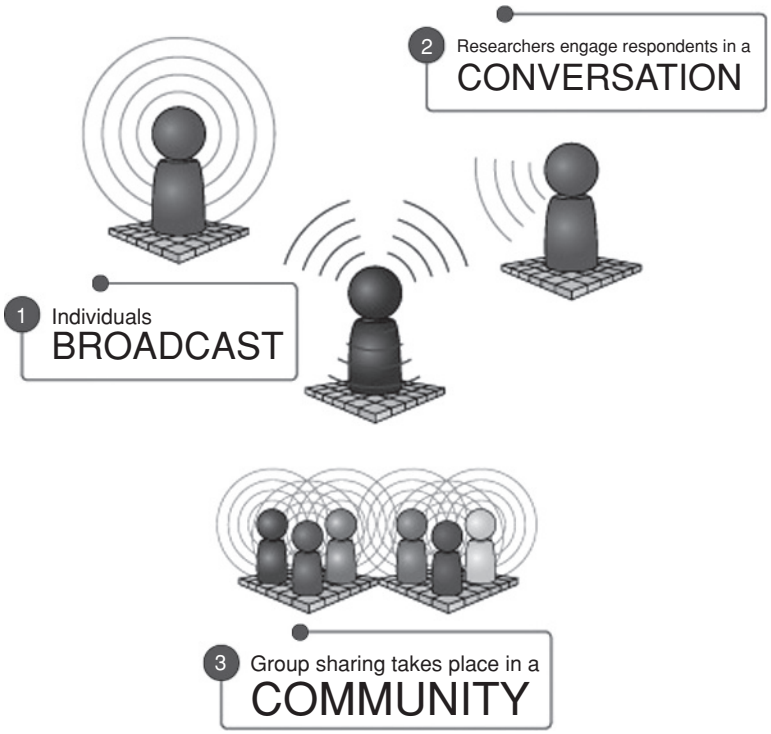


**FIGURE 1.3** The sociality hierarchy.

interest, social gaming, and forming support groups on message boards or in virtual worlds. The key concept at this communication level is *sharing* information. That is, at this level, the conversations are many-to-many—an individual may start the conversation, but many will join in, and many will "see" the conversation.

## HOW CAN SURVEY RESEARCHERS USE SOCIAL MEDIA DATA?

This volume is organized by the three levels of our sociality hierarchy. Each of the 10 pilot studies detailed in this book employs broadcast, conversational, or community-based social media or technology as an innovative approach to supplementing or enhancing survey research.

Chapters 2 and 3 examine broadcast-level social media. Twitter users broadcast their 140-character thoughts to an audience of followers—and to anyone who uses a screen-scraping tool to #discover trends and keywords in those Tweets. The organic data produced by screen-scraping can be used for nowcasting, sentiment analysis and various other analysis techniques. In Chapter 2, Haney provides a handbook for sentiment analysis, paying close attention to the pitfalls that await researchers who do not carefully consider their research question. In Chapter 3, Kim et al. perform a case study of Tweets on health-care reform to determine whether such analysis could ever replace opinion polls on the topic.

Chapters 4–7 present examples of use of the conversational properties of social media for survey research. In our model, the best use of the conversational level of the sociality hierarchy involves translating or transferring traditional survey methods to new social media platforms. In Chapter 4, Sage describes using a Facebook app to build a sample of respondents and collect survey data. In Chapter 5, Dean et al. demonstrate how the virtual world Second Life and Skype videoconferencing software can be used to conduct cognitive interviews with a geographically dispersed population. In Chapter 6, Richards and Dean use Second Life as a survey laboratory to test comprehension and compliance with the randomized response technique, a method of increasing the privacy of sensitive questions in surveys. In Chapter 7, Roe et al. describe the processes involved and decisions made in building a mobile survey panel.

Chapters 8–11 exemplify the community level of the sociality hierarchy. Community-based research takes advantage of the social and interactive elements of social media, including instant feedback (for example, sharing results with others in the group) and gamification. In Chapter 8, Keating et al. describe how crowdsourcing techniques can be used to supplement survey research. In Chapter 9, Richards et al. present a method for using Twitter to collect diary data from specific Twitter-friendly communities. In Chapter 10, Haque and Swicegood use extant social networks in Second Life to recruit and interview subjects with chronic medical conditions. Finally, in Chapter 11, Puleston describes methods for gamifying surveys—making them more interactive, interesting, and fun for respondents. In our last chapter, Hill and Dever use the sociality hierarchy to think about ways to improve the survey research of the future.

## REFERENCES

ANDERSON, P. (2007). What is Web 2.0? Ideas, technologies and implications for education. *JISC Technology and Standards Watch*. Retrieved from http://www.jisc.ac.uk/publications/reports/2007/twweb2.aspx.

ASBERG, P. (2009). Using social media in brand research. Retrieved from http://www.brandchannel.com/images/papers/433_Social_Media_Final.pdf.

BELSKY, J. (1980). Child maltreatment: An ecological integration. *American Psychologist*, *34*(4), 320–335.

BEST, S. J., & RADCLIFF, B. (2005). *Polling America: An encyclopedia of public opinion*. Westport, CT: Greenwood Press.

BHUTTA, C. B. (2012). Not by the book: Facebook as a sampling frame. *Sociological Methods & Research*, *41*(1), 57–88.

BIEMER, P. B., & LYBERG, L. A. (2003). *Introduction to survey quality.* New York: Wiley.

BIEMER, P. P., HERGET, D., MORTON, J., & WILLIS, W. G. (2000). The feasibility of monitoring field interview performance using computer audio-recorded interviewing (CARI) (pp. 1068–1073). *Proceedings of the American Statistical Association's Section on Survey Research Methods.*

BINGHAM, W., & MOORE, B. (1924). *How to interview*. New York: Harper & Row.

BLUMBERG, S. J., & LUKE, J. V. (2009). Reevaluating the need for concern regarding noncoverage bias in landline surveys. *American Journal of Public Health*, *99*(10), 1806–1810.

BLUMBERG, S. J., LUKE, J. V., GANESH, N., DAVERN, M. E., & BOUDREAUX, M. H. (2012). *Wireless substitution: State-level estimates from the National Health Interview Survey, 2010–2011*. National Center for Health Statistics. Retrieved from http://www.cdc.gov/nchs/data/nhsr/nhsr061.pdf.

BOYD, D. M., & ELLISON, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, *13*(1), article 11.

BRICK, J. M., WILLIAMS, D., & MONTAQUILA, J. M. (2011). Address-based sampling for subpopulation surveys. *Public Opinion Quarterly*, *75*(3), 409–428.

BRONFENBRENNER, U. (1994). Ecological models of human development. In *International encyclopedia of education*, Vol. 3, 2nd ed. Oxford, UK: Elsevier. Reprinted in Gauvin, M., & Cole, M. (Eds.), *Readings on the development of children,* 2nd ed. (1993, pp. 37–43) New York: Freeman.

CANNELL, C. F., & KAHN, R. (1968). Interviewing. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 2, pp. 526–595). Reading, MA: Addison-Wesley.

CARLEY-BAXTER, L., HILL, C. A., ROE, D., TWIDDY, S. E., & BAXTER, R. (2006). Changes in Response Rate Standards and Reports of Response Rate over the Past Decade. Presented at Second International Conference on Telephone Survey Methodology, Miami, FL, January 2006. Retrieved from http://www.rti.org/pubs/TSM2006_Carl-Bax_Hill_paper.pdf.

Cole-Lewis, H., & Kershaw, T. (2010). Text messaging as a tool for behavior change in disease prevention and management. *Epidemiologic Reviews*, *32*(1), 20.

Conner, M. (2012). Data on Big Data. Retrieved from http://marciaconner.com/blog/data-on-big-data/.

Coombs, L., & Freedman, R. (1964). Use of telephone interviews in a longitudinal fertility study. *Public Opinion Quarterly*, *28*(1), 112–117.

Couper, M. P. (2000). Review: Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, *64*(4), 464–494.

Cukier, K. (2010). Data, data everywhere. *The Economist*. Retrieved from http://www.economist.com/node/15557443.

Cunningham, D., Flicker, L., Murphy, J., Aldworth, W., Myers, S., & Kennet, J. (2005). *Incidence and impact of controlled access situations on nonresponse*. Presented at the annual conference of the American Association for Public Opinion Research, Miami Beach, FL.

Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly*, *69*, 87–98.

Dahlberg, L. L., & Krug, E. G. (2002). Violence-a global public health problem. In E. Krug, L. L. Dahlberg, J. A. Mercy, A. B. Zwi, & R. Lozano (Eds.), *World report on violence and health* (pp. 1–56). Geneva: World Health Organization.

de Heer, W. (1999). International response trends: Results of an international survey. *Journal of Official Statistics*, *15*(2), 129–142.

De Souza, C. S., & Preece, J. (2004). A framework for analyzing and understanding online communities. *Interacting with Computers, The Interdisciplinary Journal of Human-Computer Interaction*, *16*(3), 579–610.

Dean, E., Cook, S. L., Keating, M. D., & Murphy, J. J. (2009). Does this avatar make me look fat? Obesity & interviewing in Second Life. *Journal of Virtual Worlds Research*, *2*(2), 11.

Dean, E., Cook, S., Murphy, J., & Keating, M. (2011). The effectiveness of survey recruitment methods in Second Life. *Social Science Computer Review*, *30*(3), 324–338.

Dillman, D. (2000). *Mail and Internet surveys: The tailored design method*. New York: Wiley.

Elastic Thought (2008). Social-ecological model—the "other" SEM. Retrieved from http://elasticthought.com/2008/10/social-ecological-model—-the-other-sem.html.

Facebook (2012). Statement of rights and responsibilities. Retrieved from http://www.facebook.com/legal/terms.

Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, *99*(4), 689–723.

Goh, T., & Huang, Y. (2009). Monitoring youth depression risk in Web 2.0. *VINE*, *39*(3), 192–202.

Groves, R. M. (2011). Three eras of survey research. *Public Opinion Quarterly*, *75*(5), 861–871.

GROVES, R. M., COUPER, M. P., PRESSER, S., SINGER, E., TOURANGEAU, R., ACOSTA, G. P., & NELSON, L. (2006). Experiments in producing nonresponse bias. *Public Opinion Quarterly*, *70*(5), 720–736.

HARDY, Q. (2013). Big data in your blood. Retrieved from http://bits.blogs.nytimes.com/2012/09/07/big-data-in-your-blood/.

Health Research Institute (2012). Social media "likes" healthcare: From marketing to social business. Retrieved from http://www.pwc.com/us/en/health-industries/publications/health-care-social-media.jhtml.

HOLBROOK, A. L., GREEN, M. C., & KROSNICK, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires. *Public Opinion Quarterly*, *67*(1), 79–125.

JANSEN, B. J., ZHANG, M., SOBEL, K., & CHOWDURY, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, *60*(11), 2169–2188.

Kaiser Family Foundation (2010). Daily media use among children and teens up dramatically from five years ago. Retrieved from http://www.kff.org/entmedia/entmedia012010nr.cfm.

KEESLING, R. (2008). Controlled access. In P. Lavrakas (Ed.), *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage.

KEETER, S. (2012). Presidential address: Survey research, its new frontiers, and democracy. *Public Opinion Quarterly*, *76*(3), 600–608.

KEMPF, A. M., & REMINGTON, P. L. (2007). New challenges for telephone survey research in the twenty-first century. *Annual Review of Public Health*, *28*, 113–126.

KIM, W., JEONG, O.-R., KIM, C., & SO, J. (2010). The dark side of the Internet: Attacks, costs and responses. *Information Systems*. Retrieved from http://www.sciencedirect.com/science/article/B6V0G-51J9DS4-1/2/1a75ce9706b13898decd576f47395638.

KOZINETS, R. (2010). *Netnography: Doing ethnographic research online*. Thousand Oaks, CA: Sage.

KRISHNAMURTHY, B., & CORMODE, G. (2008). Key differences between Web 1.0 and Web 2.0. *First Monday*, *13*(6).

LAI, J. W., & SKVORETZ, J. (2011). *Utilizing Facebook application for disaster relief: Social network analysis of American Red Cross cause joiners*. Presented at the 2011 American Association for Public Opinion Research, Phoenix, AZ.

LAMENDOLA, W. (2010). Social work and social presence in an online world. *Journal of Technology in Human Services*, *28*, 108–119.

LEBERKNIGHT, C., INALTEKIN, H., CHIANG, M., & POOR, H. V. (2012). The evolution of online social networks a tutorial survey. *IEEE Signal Processing Magazine*, *29*(2), 41–52.

LEENES, R. (2010). Context is everything: Sociality and privacy in online social network sites. In M. Bezzi et al. (Eds.), *Privacy and identity*. IFIP AICT *320*, 48–65.

LINDEN, T. (2011). *2009 end of year Second Life economy wrap-up (including Q4 economy in detail)*. Retrieved from http://community.secondlife.com/t5/Features/

2009-End-of-Year-Second-Life-Economy-Wrap-up-including-Q4/ba-p/653078/page/3.

Link, M. W., Battaglia, M. P., Frankel, M. R., Osborn, L., & Mokdad, A. H. (2008). A comparison of address-based sampling (ABS) vs. random digit dialing (RDD) for general population surveys. *Public Opinion Quarterly*, *72*(1), 6–27.

Liu, H. (2007). Social network profiles as taste performances. *Journal of Computer-Mediated Communication*, *13*(1), article 13.

Liu, H., Maes, P., & Davenport, G. (2006). Unraveling the taste fabric of social networks. *International Journal on Semantic Web and Information Systems*, *2*(1).

Mazur, E. (2010). Collecting data from social networking Web sites and blogs. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 77–90). Washington, DC: American Psychological Association.

McLeroy, K. R., Bibeau, D., Steckler, A., & Glanz, K. (1988). An ecological perspective on health promotion programs. *Health Education & Behavior*, *15*(4), 351–377.

Mendelson, H. (1960). Round-table sessions current status of telephone surveys. *Public Opinion Quarterly*, *24*(3), 504–514.

Morales, L. (2011). Google and Facebook users skew young, affluent, and educated. Retrieved from http://www.gallup.com/poll/146159/facebook-google-users-skew-young-affluent-educated.aspx.

Murphy, J., Dean, E., Cook, S., & Keating, M. (2010). *The effect of interviewer image in a virtual-world survey*. RTI Press.

NielsenWire (2012). Buzz in the blogosphere: Millions more bloggers and blog readers. Retrieved from http://blog.nielsen.com/nielsenwire/online_mobile/buzz-in-the-blogosphere-millions-more-bloggers-and-blog-readers/.

O'Connell, A. (2010). Reading the public mind. *Harvard Business Review*. Retrieved from http://hbr.org/2010/10/reading-the-public-mind/ar/pr.

O'Connor, B., Balasubramanyan, R., Routledge, B., & Smith, N. (2010). From Tweets to polls: Linking text sentiment to public opinion time series. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.*

Peytchev, A. (2012). Multiple imputation for unit nonresponse and measurement error. *Public Opinion Quarterly*, *76*(2), 214–237.

Pew Internet & American Life Project (2011). Web 2.0. Retrieved from http://www.pewinternet.org/topics/Web-20.aspx?typeFilter=5.

Pew Internet & American Life Project (2012). Retrieved from http://pewinternet.org/Trend-Data-(Adults)/Internet-Adoption.aspx.

Pew Internet & American Life Project (2013). Retrieved from http://pewinternet.org/Pew-Internet-Social-Networking-full-detail.aspx.

Ravenscroft, A. (2009). Social software, Web 2.0 and learning: Status and implication of an evolving paradigm. *Journal of Computer Assisted Learning*, *25*(1), 1–5.

SCHAEFFER, N. C. 1991. Conversation with a purpose—or conversation? Interaction in the standardized interview. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys*. New York: Wiley.

SCHUTTE, M. (2009). Social software and Web 2.0 technology trends. *Online Information Review*, *33*(6), 1206–1207.

SCOTT, P. R., & JACKA, J. M. (2011). Auditing social media: A governance and risk guide. ISBN: 978-1-118-06175-6. Wiley.

SHELDON, P. (2008). The relationship between unwillingness-to-communicate and Facebook use. *Journal of Media Psychology: Theories, Methods, and Applications*, *20*(2), 67–75.

SHIN, D. H. (2010). Analysis of online social networks: A cross-national study. *Online Information Review*, *34*(3), 473–495.

SMITH, T. (2009). Conference notes-the social media revolution. *International Journal of Market Research*, *51*(4), 559–561.

STEEH, C., KIRGIS, N., CANNON, B. & DEWITT, J. (2001). Are they really as bad as they seem? Nonresponse rates at the end of the twentieth century. *Journal of Official Statistics*, *17*(2), 227–247.

SULLIVAN, S. J., SCHNEIDERS, A. G., CHEANG, C., KITTO, E., HOPIN, L., WARD, S., AHMED, O., & MCCRORY, P. (2011). What's happening? A content analysis of concussion-related traffic on Twitter. *British Journal of Sports Medicine* (published Online March 15, 2011).

SWEARER NAPOLITANO, S. M., & ESPELAGE, D. L. (2011). Expanding the social-ecological framework of bullying among youth: Lessons learned from the past and directions for the future [Chapter 1]. *Educational Psychology Papers and Publications*. Paper 140. http://digitalcommons.unl.edu/edpsychpapers/140.

The Free Dictionary (2013). Retrieved from http://www.thefreedictionary .com/sociality.

THELWALL, M. (2009). Introduction to Webometrics: Quantitative web research for the social sciences. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, *1*(1), 1–116.

THELWALL, M., WILKINSON, D., & UPPAL, S. (2010). Data mining emotion in social network communication: Gender differences in MySpace. *Journal of the American Society for Information Science and Technology*, *61*(1), 190–199.

THORNBERRY, O. T., & MASSEY, J. T. (1988). Trends in United States telephone coverage across & time and subgroups. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 25–49). New York: Wiley.

TOFFLER. A. (1980). *The third wave*. New York: Bantam.

TORAL, S. L., MARTÍNEZ-TORRES, M. R., BARRERO, F., & CORTÉS, F. (2009). An empirical study of the driving forces behind online communities, *Internet Research*, *19*(4), 378–392.

TORTORA, R. D. (2004). Response trends in a national random digit dial survey. *Metodolski zvezki*, *1*(1), 21–32.

TOURANGEAU, R. (2004). Survey research and societal change. *Annual Review of Psychology*, *55*, 775–801.

TUMASJAN, A., SPRENGER, T. O., SANDERS, P. G., & WELPE, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment (pp. 178–185). *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*.

TYNAN, D. (2002, August). Spam Inc. *PC World.*

VAN DIJCK, J. (2012). Facebook as a tool for producing sociality and connectivity. *Television & New Media*, *13*, 160.

WILKINSON, D., & THELWALL, M. (2011). Researching personal information on the public web: Methods and ethics. *Social Science Computer Review*, *29*(4), 387–401.

*Wired*. (1999, January 26). Sun on privacy: 'get over it.' Retrieved from http://www.wired.com/politics/law/news/1999/01/17538.

YEAGER, D. S., KROSNICK, J. A., CHANG, L., JAVITZ, H. S., LEVENDUSKY, M. S., SIMPSER, A., & WANG, R. (2011). Comparing the accuracy of RDD-telephone surveys and Internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly*, *75*(4), 709–747.

YUAN, Y. C., & GAY, G. (2006). Homophily of network ties and bonding and bridging social capital in computer-mediated distributed teams. *Journal of Computer-Mediated Communication*, *11*(4), 1062–1084.