6. Digital family research
Nicolas M. Legewie and Anette E. Fasang

1. INTRODUCTION

The interdisciplinary orientation of family research – spanning anthropology, history, psychology, social policy, economics, sociology, and demography – is reflected in the wide range of epistemological perspectives and methodological approaches used in the field. In some research areas, the collection of digital data has a long tradition. For instance, video data are collected to study everyday micro-level interactions in families, and are used as tools in family therapy (Erickson 2011; Reusch and Keys 1956). In other fields, and in quantitative family demography in particular, researchers increasingly employ digital traces that emerge as a by-product of other processes (Cesare et al. 2018). As digitalisation progresses, digital data are rapidly becoming abundant and accessible, and offer new and unique opportunities for conducting social research. At the same time, the quality of random probability surveys is deteriorating. The contact information that is necessary to be included in random probability sampling, for example taken from population registers, tends to be less complete and more selective with regard to the exclusion of specific social groups than it was in the past. Low response rates to telephone surveys further exacerbate the challenges entailed in traditional surveys. In the United States, for example, response rates have stabilised at around 9 per cent since 2012 following years of decline (Keeter et al. 2017). Similar trends have been observed in other countries. While the quality of many surveys is decreasing, the costs associated with conducting surveys remain higher than the costs involved in collecting most digital data (Salganik 2018). Decreasing survey quality should be of great concern to family researchers, and highlights the need to explore and improve both survey methods and other research strategies, including administrative records and digital traces.

This chapter focuses on the opportunities and challenges associated with using different types of digital data in empirical family research. We distinguish between 1) digital traces that are repurposed for family research, including data from Twitter, Facebook, or Google searches; and 2) digital data collection tools developed and used specifically for research purposes, particularly video data analysis and experience sampling. We build on Salganik’s (2018, p. 7) distinction between digital traces as ‘ready-made’ data, and digital data that have been collected for research purposes as ‘custom-made’ data. In practice, researchers often choose strategies that combine ready-made and custom-made digital data, or employ additional sample selection procedures that ‘customise’ ready-made digital traces to some degree. While both types of digital data can be used to address a range of research questions, there are a number of challenges associated with collecting and analysing these data.

Here, we will not discuss the question of how digital technologies, such as smartphones or video calls, affect family life, as this issue can be addressed using digital data, conventional survey data, qualitative interviews, ethnographic methods, or other research approaches. Instead, we focus on the opportunities for using these two types of digital data in family research more generally. Rather than attempting to provide a comprehensive review of the
studies on family issues that have used digital data, we will present examples of studies that shed light on the areas in which the application of digital data in family research seems particularly promising.

We begin by introducing ready-made digital trace data (Section 2), and comparing these data to custom-made data purposefully collected with digital technologies (Section 3). We explore the advantages and challenges associated with using each type of data, and present examples of how these data are applied in family research. Moreover, we seek to identify family-related research questions for which the use of ready-made and custom-made digital data appears particularly promising. In Section 4, we close by providing a broader perspective on the opportunities and pitfalls associated with using digital data in family research, including reflections on the ethics of conducting such research.

2. DIGITAL TRACE DATA

Digital traces, or ‘digital footprints’, result from ‘social interaction via digital tools and spaces as well as digital records of other culturally relevant materials, such as archived newspapers and Google searches’ (Manovich 2011). These traces include data from popular social networking sites (such as Facebook or Twitter), personal blogs, and collaborative online spaces (such as Wikipedia), as well as data derived from mobile phone or credit card usage (Cesare et al. 2018). Thus, a defining characteristic of digital traces is that they emerge as by-products of social processes. Hence, digital traces can be seen as ‘ready-made’ datasets that were not gathered primarily for research purposes (Salganik 2018).

The use of digital trace data is at the core of a rapidly emerging field of ‘digital demography’ (Alburez-Gutierrez et al. 2019; Blumenstock 2018; Blumenstock et al. 2018; Cesare et al. 2018). To date, digital demography has mostly focused on analysing international migration flows, including the location tracking of IP addresses when individuals log into their email accounts, Facebook advertisement target populations, geo-located Twitter data, and Google+ data (Zagheni and Weber 2012). Blumenstock et al. (2015, 2018) have called attention to both the possibilities and the caveats that come with using cell phone data to influence and to study global socio-economic development, particularly in poorer regions of the world.

While migration and global development are undoubtedly tightly linked to processes unfolding within family networks, the use of digital trace data to study core topics in family research has been relatively limited. The area of family research in which digital trace data have been used most extensively, and for the longest period of time, appears to be research on online dating (Hitsch et al. 2010). Online dating as a means of mate selection is becoming increasingly accepted and widespread. A large number of studies have used data from online dating websites to draw conclusions about mate selection, and about decision-making more generally (Bruch and Newman 2018; Skopek et al. 2011; Van Bavel in this volume). The findings of online dating studies have substantiated previous research showing that there is considerable educational homogamy in partner selection on online dating platforms, affecting how individuals both contact and respond to prospective matches, albeit in gender-specific ways. It has, for example, been shown that higher-educated men tend to contact lower- as well as higher-educated women, whereas higher-educated women are less likely to contact or reply to lower-educated men (Skopek et al. 2011). In addition, the findings of online dating studies have shed light on different stages of the decision-making processes in mate selection, which
Digital family research could have more general implications for decision-making behaviour in market situations (Bruch et al. 2016).

Several recent studies have started to explore digital trace data for predicting fertility trends. These studies have, for example, used web searches to show that searches of terms such as ‘abortion’, ‘pregnancy’, or ‘birth’ predict short-term future behaviour, at least on the level of aggregated averages (Ojala et al. 2017; Rampazzo et al. 2018). Ojala et al. (2017) combined digital traces from Google Correlate and Google Trends on fertility with custom-made ‘ground truth data’ from the American Community Survey to show that specific searches vary with socio-economic status in expected ways. In addition, recent studies are beginning to transfer web-scraping methods and natural language-processing techniques that were developed in computer sciences, and are frequently applied in political science to family policies and public discourses around family issues. For example, Gülzau (2018) used topic modelling to link digitalised media coverage of family issues to family policy change, and analysed exchanges in a parenting online forum to study how parents negotiate parenting norms.

Ready-made video data represent another potentially interesting source for family researchers. User-generated content websites are providing an ever expanding pool of visual data (i.e., moving or still images) that are easily accessible for researchers (Nassauer and Legewie 2018). Most of the existing studies that have used these data addressed the impact of online platforms such as YouTube, Instagram, or Facebook on family interactions in combination with custom-made observational and survey methods (e.g., Baker and Carreño 2016; Burroughs 2017; Charoensukmongkol 2018; Fingerman et al. 2016; Jeanfreau et al. 2019; Le Moignan et al. 2017; Racz et al. 2017). But data from these platforms could be used much more broadly, as they offer direct glimpses into everyday family interactions. While this opportunity is being increasingly recognised in other fields, such as criminology and deviant behaviour research (e.g., Collins 2008; Lindegaard et al. 2018; Nassauer 2016), family researchers who take advantage of ready-made video data remain the exception. For instance, Marcon et al. (2019) studied breastfeeding displays on Instagram, and examined how users share perspectives and information about the topic. Johnson et al. (2019) used YouTube videos to study the lived experiences of parents and families affected by the opioid crisis. Researchers can access platforms such as Facebook and Instagram to study videos and pictures as well as the captions that users assign them, and comments from their social network that the videos and pictures elicit. For instance, such data could be used to study (online) presentations of the self during family formation, or norms of parenting and relationship harmony.

Thus, while researchers have made extensive use of online data to study dating, they have only started to explore opportunities to draw upon online data to study topics such as fertility, family policy, and parenting norms and practices. This reluctance could be attributable to the difficult trade-offs involved in using such data. It may, for example, be hard to interpret the meaning of what is being measured in the digital traces (see below).

2.1 Opportunities

Digital traces have been heralded as a third major data revolution in ‘digital demography’, after the paradigms of ‘census and administrative records’ and ‘theory-driven micro-level data’ (Billari and Zagheni 2017). Using digital trace data has several key advantages over using conventional survey data, especially when seeking to answer specific questions in family research (Mützel 2019; Salganik 2018; see also for an overview Table 6.1).
### Table 6.1 Overview of the possibilities and the challenges associated with using ready-made and custom-made digital data for family research

<table>
<thead>
<tr>
<th></th>
<th>Digital traces: ready-made</th>
<th>Digitally collected: custom-made (smartphone-based experience sampling methods and video-based research)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Potential</strong></td>
<td>– Big, large N</td>
<td>– Data collection of real-time, or close to real-time actions and situations</td>
</tr>
<tr>
<td></td>
<td>– Always on, real time, often repeated measurements in small intervals</td>
<td>– Detailed study of micro-level processes</td>
</tr>
<tr>
<td></td>
<td>– Non-reactive, no social desirability bias towards researchers</td>
<td></td>
</tr>
<tr>
<td><strong>Challenges</strong></td>
<td>– Limited information on the population covered: selective, non-representative, drifting, algorithmically confounded (bots)</td>
<td>– Limited information on individuals and the context: on their own, neither experience sampling methods nor video data analysis provide detailed information</td>
</tr>
<tr>
<td></td>
<td>– Limited information on data-generating process: data dirty or inaccessible, often held by private companies, data incomplete to assess theoretical constructs, indicators available have ambiguous interpretation</td>
<td>– New skill set requirements for researchers: e.g., building an app and data infrastructure or methods of video data analysis</td>
</tr>
<tr>
<td></td>
<td>– Ethically sensitive: lack of informed consent, data often owned by private companies, potential for misuse</td>
<td>– Information on study subjects and context</td>
</tr>
<tr>
<td></td>
<td>– Ethically sensitive: lack of informed consent, privacy, potentially disruptive Hawthorne effects</td>
<td>– Ethically sensitive: lack of informed consent, privacy, potentially disruptive Hawthorne effects</td>
</tr>
<tr>
<td><strong>Potential solutions</strong></td>
<td>– Extensive data cleaning</td>
<td>– Further training, methodological development, and interdisciplinary collaboration</td>
</tr>
<tr>
<td></td>
<td>– Check against ground truth data</td>
<td>– Mixed-methods approaches</td>
</tr>
<tr>
<td></td>
<td>– Combination with other data sources</td>
<td>– Careful assessment of research ethics, walled-garden data archiving, diligent pre-study briefing of participants</td>
</tr>
<tr>
<td></td>
<td>– Post-stratification</td>
<td></td>
</tr>
<tr>
<td><strong>Applications</strong></td>
<td>– Online dating</td>
<td>– Division of housework</td>
</tr>
<tr>
<td></td>
<td>– Google searches on fertility</td>
<td>– Families as learning environments</td>
</tr>
<tr>
<td></td>
<td>– Parenting online forum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>– Digitalised newspaper coverage of family issues</td>
<td></td>
</tr>
<tr>
<td><strong>Promising areas of future application</strong></td>
<td>– Small subpopulations that are difficult to reach and/or tend to be under-covered in surveys: male fertility, LGBTIQ population</td>
<td>– Family interactions</td>
</tr>
<tr>
<td></td>
<td>– Self-representation, family norms</td>
<td>– Situational dynamics and micro-level processes</td>
</tr>
<tr>
<td></td>
<td>– Family time use, division of labour</td>
<td>– Family time use, division of labour</td>
</tr>
<tr>
<td></td>
<td>– Group comparisons rather than representative population estimates</td>
<td>– Self-representation, family norms</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Resource activation in family networks</td>
</tr>
</tbody>
</table>

**Source:** Cesare et al. (2018); Mützel (2019); Nassauer and Legewie (2018); Salganik (2018).

**Note:** a Lesbian, Gay, Bisexual, Transsexual, Intersexual, Queer.

First, these datasets are usually big. Thus, family researchers using digital trace data are less limited by sample size restrictions; a problem that often arises when researchers attempt to use survey data to study specific subpopulations with small case numbers, such as single fathers, fathers who take parental leave, female breadwinner families, or families with special needs. For example, even when using exceptionally rich survey datasets, such as the German Family Panel pairfam (Huink et al. 2011), researchers quickly run into sample size limitations due to attrition and selective non-response. But it is also worth noting that some subpopulations are relatively well covered on social media and the internet, while others are not. For example,
members of the LGBTQ community tend to be active on dating websites and in social media, because these relatively anonymous virtual forums allow them to interact more effectively as members of a relatively small and geographically dispersed population. In contrast, young children are not covered very well in social media, except through the eyes of a select group of parents who share content about their children online, despite the known risks of the misuse of such material, particularly if it is visual. The use of digital trace data is particularly promising for quantitative studies of subpopulations who are usually not well covered in surveys, and who are often only analysed using small, qualitative snowball or convenience samples (Fasang et al. 2016). These data can be used to examine regional variation more effectively than has been previously possible by enabling researchers to study local institutional and compositional context effects on small subpopulations. For example, social media data could be employed to map the local concentrations of specific subpopulations more accurately, and to investigate how the local concentrations of these groups affect their interactions and outcomes.

Second, as most digital traces are ‘always on’, they can provide real-time repeated measurements in very short intervals (Lazer and Radford 2017; Salganik 2018). Thus, by drawing on these data, researchers can explore the temporal dynamics of behaviour and interactions in much greater detail than is possible using survey data. Moreover, the continuous production of digital traces is ideal for research designs that build on natural experiments with before and after comparisons. As a caveat to these advantages, the population of users covered in the processes that generate digital trace data, and the search and newsfeed algorithms that digital platforms employ (e.g., Google search, Facebook, or Twitter; Lazer et al. 2014), are constantly shifting and changing; a process Salganik (2018) refers to as ‘drifting’ (see below). Therefore, these data are not well suited for detecting changes over time or in an aggregate-level process, as any changes that are found might simply be due to drift in the user population or algorithms that cannot easily be monitored by the researcher. Drifting severely limits the options researchers have for using digital trace data to study social change in family structures at the aggregate level. Nonetheless, family researchers examining micro-level family dynamics may find it interesting to study the temporality of short-term interactions among, for example, different generations on Facebook. It may be assumed that within short time intervals, the population covered in most digital traces remains reasonably stable.

Third, digital traces are ‘non-reactive’. This means that because these traces are a by-product of other processes, the possibility that participants adapted their actions in response to being observed is low. The non-reactivity of digital trace data could be particularly appealing for addressing family-related questions, which are often complicated by issues of social desirability bias. Many topics related to family, fertility, and intimate relationships are normatively loaded, such as abortion or extra-marital relationships. However, this potential advantage may not be as clear-cut as it first appears to be. While digital traces are, by definition, not affected by social desirability bias towards the researcher, the self-representation in social media will be heavily guided by social desirability in the general population. For instance, research has shown systematic bias in online dating platforms, with men over-reporting their height and women under-reporting their weight (Toma et al. 2008). Similarly, the representation of family life on social media platforms often polarises into the ‘showing off’ of happy and successful family events, interactions, and achievements on the one hand; and the rejection of unrealistic norms of parenting and relationship harmony through the sharing of stories about the failure to meet these goals by self-help-oriented groups on the other (Gülzau 2018). Hence, digital traces may be more likely to provide information about family norms and how
they are negotiated than to give ‘true’ reports about family behaviour and interactions. For these reasons, it is crucial that family researchers using these data distinguish digital traces that reflect a conscious representation of self from digital traces that indicate ‘true’ behaviour and characteristics (see our discussion of the ambiguity of the social meaning of digital traces below). Researchers may be able to discern actual behaviour through a strategy of ‘partnering with the powerful’ (Salganik 2018); i.e., collaborating with organisations such as non-governmental organisations, companies, and app developers who generate digital traces. For example, anonymised data from family calendar apps such as Cozi, CloudCal, or Looping could be helpful in studying family time use and the gender division of paid and unpaid work. It is well documented that individuals tend to over-report the share of the housework they do, but under-report the share of the housework their partner does. In surveys, partners’ assessments of their own and each other’s contributions to housework often deviate considerably, and the use of different measures leads to different results. Lee and Waite (2005) highlighted this issue by benchmarking survey questions against estimates obtained from an experience sampling study (see below). While actual family behaviour will still deviate from plans fixed in a calendar, these plans tend to be more accurate than individual self-reports, and are more likely to reflect the outcomes of family negotiations on the planned division of tasks. Of course, the need for ethical safeguards and effective anonymisation can make conducting this type of research especially challenging (Salganik 2018, p. 307ff.).

2.2 Challenges

While there is considerable excitement surrounding digital demography, there are many challenges involved in repurposing digital traces for social research in general, and family research specifically. The rather extensive list of potential challenges mentioned in the literature includes issues related to coverage, selectivity, and non-representativeness; lack of information on data-generating processes; and ethical questions regarding the collection of, access to, and analysis of digital trace data (Cesare et al. 2018; Fasang 2018; Legewie and Nassauer 2018; Mützel 2019; Salganik 2018). These potential challenges have been conceptually organised in different ways. We have chosen to group them around three main topics: limited information on the population covered, limited information on the data-generating processes, and ethical sensitivity.

2.2.1 Limited information on the population covered

Most digital trace data completely cover a subset of users of a particular online service. As a result, the sample is often highly selective, and is rarely representative of a general population that family researchers are interested in. These biases create challenges akin to biases in other types of data, such as selective non-response in surveys, or sampling on the dependent variable in qualitative studies (for a discussion of this issue, see, e.g., Lazer and Radford 2017, p. 31). As in the case of survey data, the selectivity and non-representativeness of digital data can be quantified and modelled with appropriate statistical procedures, and countered with post-stratification measures. For example, weighting procedures allow accounting for bias based on known proportions in the population of interest (Salganik 2018). Billari and Zagheni (2017, p. 9) consider social media and the internet as ‘laboratories’ that produce systematically biased estimates of quantities: ‘there are hidden, potentially stochastic rules that determine the relationship between the online data and the offline quantities of interest’.
Bias can then be modelled against ground truth data. If ground truth data do not exist, different hypothetical scenarios can be developed as plausible upper and lower bounds of estimates. Given that there are applications using digital data with elaborate post-stratification methods in other fields of study, such as voting (Wang et al. 2015), we are optimistic that the selectivity and non-representativeness of digital trace data will be manageable in many applications of interest to family researchers.

Drifting – i.e., the tendency of user populations and algorithms underlying online platforms to change in unknown ways over time – further limits the potential application of digital data to family-related research questions. For example, it is arguably the case that questions regarding changes in population-level family structures will continue to be better addressed using representative census or survey data, or using vital and residential register data, if they are available. Drifting also creates challenges for the post-stratification of digital traces when addressing longitudinal research questions. It is likely that accurate post-stratification will be far easier within a brief time window than over a longer time span in which compositional shifts in the user population or changes to the data-generating algorithm may have occurred.

Finally, the population covered in online platforms also tends to include bots that create systematic noise in the data through fake identities. The presence of such bots makes digital data ‘dirty’, and has led to serious misinterpretations of content (Salganik 2018), as well as to calls for extensive data cleaning (Mützel 2019). While statistical post-stratification can be done using known distributions in the population, data cleaning requires researchers to develop a different set of skills. For example, before analysing digital traces, researchers may want to identify and delete information from fake identities. Successfully performing both post-stratification and the data cleaning of digital traces relies on having an understanding of the social processes that create the biases. Therefore, researchers should gather as much information as possible, e.g., on the influence of bots in the digital trace data used, to address a specific research question.

For applied family researchers, there are two important points regarding the limited information on the population covered. First, many family-related researchers are interested in comparing groups for whom they do not need representative data at the population level. For addressing these types of questions, digital traces might be particularly well suited and relatively easy to analyse, although extensive data cleaning will still be necessary in most cases. As the skills needed for cleaning digital data differ from those needed for cleaning survey data, researchers will benefit from having opportunities to acquire both sets of skills (Mützel 2019). Second, addressing questions based on representative estimates at the population level will require considerable investments in learning how to combine post-stratification with other data sources. In the years to come, data cleaning and post-stratification skills will have to be actively developed in the digital social sciences (Salganik 2018).

2.2.2 Limited information on the data-generating processes

Unlike other forms of data researchers use, digital data are generated through the large-scale involvement of commercial companies, government agencies, as well as other organizations. In many cases, the information needed to understand how certain data were generated is not recorded or is not accessible to the researcher, and the underlying data production processes may change over time without the researcher’s knowledge (Lazer et al. 2014). Digital trace data might reflect the baked-in strategic decisions made by the platform designers rather than the natural behaviour of individuals; a process Salganik (2018) calls algorithmic confounding.
For example, Facebook encourages individuals to make more friends until they have reached 20 friends, and then stops. As a result, an unusually high proportion of Facebook users have exactly 20 friends, which would also bias the representation of family networks in Facebook data (see Salganik 2018 for an elaboration of this example). The design of online platforms, including their strategic interests, shape the data in ways that will often bias the estimates family scholars are interested in. In addition, digital data usually do not provide all of the information researchers need in order to capture their theoretical constructs. As a result of algorithmic confounding and incompleteness, the meaning of digital trace data is often ambiguous. What do ‘likes’ mean, and how do their meanings vary in different online contexts, or when a platform changes (e.g., when Twitter changes options for re-tweeting or Facebook adds a ‘dislike’ button)? In terms of custom-made videos and other graphical materials, it is often unclear who uploads what as a video or a picture to YouTube or Instagram, which uploads the platform’s search engine show on a given device, and how the search algorithm changes over time (Legewie 2019; Legewie et al. 2019).

Applied family researchers might deal with these sources of bias in different ways. First, it is important to have as complete an understanding as possible of the platforms that are generating the digital trace data. What are the strategic interests of the organizations involved? Which algorithmic processes influence the collection of the digital traces? Were changes to the algorithm implemented during data collection? If too little information on these processes is available, or too much cofounding is detected, then repurposing this particular type of digital trace for use in social research may not be possible (Salganik 2018). It may, however, be feasible to conduct small-scale studies that combine survey, qualitative, and even ethnographic data on the users of online platforms in order to contextualise the meaning of incomplete digital trace data.

2.2.3 Ethical sensitivity

Another key issue that arises in repurposing digital trace data for social science research is how to deal with the ethical challenges. The term ‘digital traces’ was first introduced by Latour (2007), and refers to the fact that digitalisation has increased the general traceability of human behaviour and interactions. ‘It is as if the inner workings of private worlds have been pried open because their inputs and outputs have become thoroughly traceable’ (Latour 2007, p. 2). This development can affect four basic principles of research ethics: beneficence, which means that a study’s benefits should outweigh its risks, and that researchers should minimise the risks involved for participants or research subjects (RATSWD 2017, p. 10; Sumner, 2006, p. 96); respect for persons, that is, the acknowledgement of peoples’ rights to self-determination, personality, and privacy (Gebel et al. 2015: §8–9; Salganik 2018, p. 295); justice, which calls for a fair distribution of a study’s risks and benefits (Salganik 2018, p. 298); and respect for law and public interest, which refers to the importance of complying with existing law and ensuring transparency-based accountability in the research process (Salganik 2018, p. 300).

In contrast to respondents in survey and census data, users of online services have often not given explicit consent for their data to be used for research purposes, or at least are not aware that they have given consent. Research findings based on digital trace data can be highly sensitive. For instance, digital traces may be used to quantify local populations of vulnerable groups, include information on minors, or contain information on behaviour and actions that may be subject to prosecution by authorities. It took decades for the potential for human rights abuses based on census and survey data to become fully apparent, including genocide and
forced migration (Seltzer and Anderson 2001). There are many well-documented historical cases of devastating forms of population data misuse, which have triggered an intense discussion about potential safeguards (Seltzer and Anderson 2001). Moreover, according to recent studies, it is surprisingly easy to de-anonymise digital data, particularly once they are linked with other types of data (Salganik 2018). At the same time, the opportunity for data linkage greatly increases the potential pay-off associated with using digital traces. Against this background, rigorous adherence to research ethics and the development of effective de-anonymisation techniques have to be at the core of the field of digital social science, particularly given the massive involvement of private companies in collecting these data. Recent scandals involving Facebook are raising awareness about ethical issues around digital trace data, which are already the focus of emerging research (Cesare et al. 2018; Salganik 2018; Zimmer 2010).

3. CUSTOM-MADE DIGITAL DATA

Custom-made digital data are data that were purposefully collected by researchers to answer a specific question. These data are distinct from other types of custom-made data, such as data from classic surveys or in-depth interviews, in that digital tools were used to collect them (Salganik 2018, p. 7). Established research methods for purposeful data collection, such as in-depth interviewing, ethnographic field work, and survey research, have produced countless intriguing findings for family research, including on topics such as parenting styles and practices (Lareau 2003; Spera 2005; Steinberg et al. 1992), organisation of household and reproductive labour (Bianchi et al. 2000; Hook 2010), negotiation of gender roles (Bielby and Bielby 1992; Daly 2002; Sassler and Miller 2011), intergenerational relations (Bengtson and Troll 1978; Bengtson et al. 2002), and families as learning environments (Bäumer et al. 2011; Marjoribanks 2019). Custom-made digital data and research methods may be employed to complement these established methods, and could offer additional insights. The approaches that may be used range from digital surveys (Adolph et al. 2017; Runyan et al. 2013; Salganik and Levy 2015), to video data collection (Nassauer and Legewie 2018, 2019; Elsner and Wertz 2019), to digital field experiments and mass collaborative data collection (Bail et al. 2018; Cooper et al. 2010). Here, we will focus on two examples: digital surveys conducted with smartphone-based experience sampling methods (ESM), and video data collection.

ESM (Hektner et al. 2007; see Stone and Shiffman 1994 for a similar approach) were developed to investigate people’s subjective experiences in situ, and the use of these methods is well established in research on physical and mental health, happiness, and emotional states (e.g., Brose et al. 2017; Cohn et al. 2011; Kimhy et al. 2006). In family research, ESM have been employed to study issues such as emotional transmission in families’ daily lives (Larson and Almeida 1999), family communication and time use and their connection to adolescents’ well-being (Offer 2013a, 2013b), and convergence of parents’ and adolescents’ emotional states (Larson and Richards 1994). But there are also ESM applications that have been used to study not just personal experiences, but social actions and processes (Zirkel et al. 2015, p. 7). In a pioneering application in family research, Lee and Waite (2005) studied the division of housework in married couples. Their findings indicated that husbands over-estimate their own contributions to household work, that couples have different perceptions of the gap in the household work performed by the wife and the husband, and that survey studies may
over-estimate overall time spent on household work. Such applications suggest that there is potential for a much broader use of smartphone-based ESM in family research.

Regarding video data collection, some studies have used new technologies to collect self-recorded video data. In an early example, researchers from the University of California Los Angeles Sloan Center on Everyday Lives of Families video-recorded families’ morning, evening, and weekend routines (Ochs and Kremer-Sadlik 2013). Recently, the New Jersey Families Study placed auto-activated video cameras in up to four rooms in participants’ homes for two weeks to study how families build skills in their pre-school children. In the field of developmental psychology, researchers use video recordings to study early childhood development and the family as a learning environment. Going beyond reliance on traditional approaches, researchers are increasingly using either custom-made or ready-made video data from real-life settings to complement laboratory experiments (Elsner and Wertz 2019).

3.1 Opportunities

The use of ESM and video data collection illustrates that there are a range of opportunities for applying custom-made digital data. Smartphone-based ESM ‘enable us to learn about individuals’ lives in context by measuring participants’ feelings, thoughts, actions, context, and/or activities as they go about their daily lives’ (Zirkel et al. 2015, p. 7). ESM are used to collect data by repeatedly sending study participants micro-surveys of only a few (closed or open) questions each that are triggered by a certain event, based on a time schedule, or generated at random over the duration of the study. Thus, ESM studies can avoid asking retrospective and generalised questions, and are instead able to collect data close to a situation or a process of interest, and to provide information on intra-individual variance through repeated measures (Hektner et al. 2007; Runyan et al. 2013; Salganik 2018; Zirkel et al. 2015). Hence, smartphone-based ESM have the potential to complement traditional methods in a number of ways. First, these methods address issues of ecological validity and retrospective generalised responses. ESM can be used to collect data on personal experiences, actions, and processes at various points in time and across environments throughout the course of a day, all in real (or close to real) time. As a consequence, smartphone-based ESM are better able to capture social processes and to provide more valid data than retrospective data from surveys or in-depth retrospective interviews. Moreover, compared with traditional surveys, ESM can provide a more nuanced, intra-personal view on a given phenomenon. For instance, in family research, smartphone-based data collection could be used to examine the relatively under-studied issue of the dynamics of resource activation in family networks (Dika and Singh 2002, p. 42; Portes 1998). Over the course of a set period, family members could be asked questions in the form of micro-surveys about situations in which they required help; transmissions of instrumental and emotional support; from which family members they received this support; whether they asked for help or received help without asking; and their experiences of these situations. The micro-surveys could be administered several times a week, yielding detailed information on personal network dynamics.

The opportunities for using video data collection in family research have increased dramatically since the early 2000s. Video equipment has become so affordable and relatively unobtrusive that conducting studies in which video data are collected directly in families’ homes is becoming a viable option. Such videos capture the sequential nature of interactions, and because they provide extremely detailed information, they allow researchers to trace the
digital family research

Researchers can use videos to study behaviour, interactions, and situational dynamics and their impact on the social outcomes (e.g., Nassauer and Legewie 2018, 2019) and patterns in communication (e.g., Norris 2004) that are central for understanding dynamics in families. For instance, video data could be used to study situational negotiations of gender roles or the division of household labour; what norms or narratives are invoked in such situations; what emotional dynamics are at play during negotiations; and how negotiations play out within the larger context of the family’s and each individual’s daily routines.

In principle, both ready-made and custom-made video data could be used to study family interactions that either take place in a given space, such as the home, or that revolve around a specific type of event, such as a wedding, a family dinner, or even a large-scale public event. One of the biggest advantages of using video data rather than participant observations and in-depth retrospective interviews is that such data provide situational information with unprecedented levels of detail and reliability (Nassauer and Legewie 2018, pp. 6–7). Video data enable researchers to conduct in-depth, second-by-second analyses of social situations, as they offer the option of rewinding the video or watching it in slow motion to focus on even fleeting pieces of information (such as facial expressions); or of reconstructing the exact sequence of interactions frame by frame, even during long and complex situations (Nassauer and Legewie 2018, p. 8). The use of video data also increases the reliability of findings because multiple researchers can analyse the same raw data material (Nassauer and Legewie 2018, p. 8). For instance, in addition to being employed in analyses of the family as a learning environment, the New Jersey Families Study data mentioned above could be used to study negotiations around reproductive labour in families. Situational negotiations could be analysed in great detail, with verbal content, body language, and the use of space and symbolic actions being examined to gain a more detailed picture than would be possible using survey data, in-depth interview data, or even participant observation data.

3.2 Challenges

Overall, ready-made and custom-made digital data offer researchers rather distinct complementary opportunities (see Table 6.1). The challenges that arise for researchers when using these data overlap considerably, and are in part the same challenges that arise when using other types of data. As custom-made digital data may be collected with a broad range of methods, their use may entail a diverse set of challenges. Here, we focus on three challenges researchers commonly face when using both ESM and video data collection approaches.

3.2.1 Limited information on individuals and the context covered

Smartphone-based ESM and video-based research may provide limited information on the individuals covered and the context of the behaviour and interactions, albeit in different ways. Smartphone-based ESM are not geared towards administering long surveys, which limits the amount of background information researchers are able to collect using these methods. Video-based research provides less tacit knowledge than participant observation. Being physically present in a social environment over long periods of time, and being able to absorb that environment with all human senses, can provide a fuller picture than video data alone (Nassauer and Legewie 2018, pp. 25–6). Moreover, such data may not supply enough background information about individuals and the physical and social context in which
a situation takes place. Mixed-methods approaches can help to overcome these challenges. Smartphone-based ESM studies can be integrated into existing in-depth interview studies or cross-sectional or panel surveys that provide relevant background information on individuals. For video-based studies, researchers may consider conducting surveys with participants if they collect video data themselves (as researchers did for the New Jersey Families Study), contacting uploaders of online videos to interview them, or collecting additional non-video data online (Nassauer and Legewie 2020).

3.2.2 New skill set requirements for researchers
Smartphone-based data collection may require building an app and an infrastructure to collect and store the data; processes with which many social scientists may be unfamiliar. Moreover, while this method has been tested, and is an established tool that is mainly used for collecting data on emotional states, it has been used much less frequently to study processes such as personal network dynamics. Thus, when using this approach, researchers may need to develop new ways of improving data quality and reducing total survey error (i.e., coverage error, unit and item non-response, measurement error; see Biemer et al. 2017; Groves and Lyberg 2010). In terms of video-based research, most researchers are not trained in, and are largely unfamiliar with, the use of video data (for an introduction, see Nassauer and Legewie 2018, 2019; Norris 2004; among many others). In addition, since video recording devices have been widely available for a much shorter period of time than voice recording devices, video methods still require a great deal of development. The use of custom-made online video data can present further challenges. Like the new techniques necessary for cleaning ready-made data, these are challenging issues in the current context, but the potential solutions are relatively straightforward. Interdisciplinary collaboration can help researchers conduct studies for which they lack certain technical skills, and further methodological development and training can empower researchers to tackle some of the open methodological questions, and to familiarise themselves with the methods.

3.2.3 Ethical sensitivity
The use of smartphone-based ESM has been shown to carry a considerable risk of reactivity; that is, of participants adapting their behaviour because of the study. While this tendency can be desirable, and even suggests opportunities for conducting intervention studies (Runyan et al. 2013), it may also have unintended consequences, disrupt participants’ lives, and undermine the quality of the data. Using video data may present research ethics challenges, mainly regarding informed consent. When filming in a family home, such as in the New Jersey Families Study, it becomes a challenge to ensure that everyone shown in a given video consented to being filmed (but see Golann, Mirakhur and Espenshade 2019 for how to handle this challenge). The same is true for online video data, and in addition people may not have agreed to that video being uploaded online (see Section 2.2.3; Legewie and Nassauer 2018, p. 6ff.). To tackle this set of challenges, researchers must engage in careful preparation and reflection. When applying ESM, it is crucial to ensure that the participants are briefed in detail about what their participation may entail, including the possibility of fostering reflection on and changes to their behaviour (see Sugie 2018 for an exemplary approach to participant briefing). When performing video-based studies, researchers should reflect on the type of video they are using and the context from which the data are being collected, and weigh the advantages of these approaches against other dimensions of research ethics, such as the unique opportunities and
the potential harm associated with the research (for a guideline, see Legewie and Nassauer 2018, 2020). Depending on the specific research context (such as the type of behaviour under study or whether faces are visible in the videos), lack of informed consent may be evaluated as more or less problematic.

4. OUTLOOK

In response to the digital data revolution and the declining quality of costly conventional survey research (Keeter et al. 2017; Salganik 2018), social scientists are increasingly exploring newly available sources of digital data, and techniques for analysing these data. New digital data collection tools developed for research purposes are often far more cost effective than conventional approaches, and can open up entirely new lines of inquiry. While the use of digital data will certainly not provide answers to all questions in family research, scholars in this field should attempt to identify the areas in which digital data can be employed most effectively to complement the use of conventional data sources and data collection strategies. In this chapter, we discussed the opportunities as well as the methodological and ethical challenges associated with using two different types of digital data to investigate family-related research questions: 1) ready-made digital data, i.e., digital traces that emerge as a by-product of other processes, and are repurposed for family research; and 2) custom-made digital data, i.e., digital data that are purposefully collected with new digital data collection tools in order to conduct family research. We presented the opportunities, the challenges, and the areas of application in family research associated with both types of digital data. While using ready-made and custom-made digital data offers complementary opportunities, the challenges involved in using these data overlap. Among the issues that can arise in using these data are having limited information to contextualise the meaning of the digital data, needing to develop new skill sets for collecting and analysing these types of data, and ensuring the ethically responsible use of these data for research purposes.

We conclude by observing that the use of digital trace data is particularly promising for addressing research questions for which other data do not exist or are difficult to collect, and when combined with ‘traditional’ forms of data from surveys and administrative records (Billari and Zagheni 2017). The study of digital trace data may prove especially helpful, as such data can be used to study small subpopulations that are difficult to reach and/or tend to be under-represented in surveys (e.g., single fathers or LGBTIQ community); self-representation and family norms; family time use and division of labour; and group comparisons (rather than representative population estimates in family research). Post-stratification, data cleaning, and effective anonymisation, as well as securing the active consent of participants are the core areas in which advancements are needed to enable researchers to fully capitalise on the potential of digital trace data.

In examining the use of custom-made digital data, we focused on smartphone-based ESM and video data collection. These methods are particularly promising if researchers are attempting to study real-life actions, interactions, and situations in the family to gain detailed insights into micro-level processes. Setting up a series of tailor-made micro-surveys delivered to respondents’ smartphones throughout the day, or collecting videos of real-life family interactions in situ, can provide researchers with unprecedented glimpses into everyday family dynamics. These studies may include questions on a range of topics, including on family inter-
actions, role negotiations, and resource activation in family networks. ESM and video data can also be used to complement digital trace data when studying family time use and division of labour, as well as self-representation and family norms.

Both ready-made digital traces and custom-made digital data can be analysed using quantitative or qualitative methods. The qualitative content of big data sources such as Facebook and Twitter posts – which has received much more attention in political science than in demography or family research – could prove particularly valuable for theory development, as selectivity and non-representativeness are less problematic in this context (Fasang 2018). By drawing on the advances and innovations in digital research methods that have been made in recent years, family researchers stand to benefit from new opportunities and ways to tackle some of the core questions in the field.

ACKNOWLEDGEMENTS

Anette E. Fasang gratefully acknowledges funding from the Swedish Research Council for Health, Working Life and Welfare (FORTE, grant no 2018–01612) for supporting a research visit at the Stockholm Demography Unit (SUDA) in spring 2019, during which part of this chapter was written.

Nicolas M. Legewie gratefully acknowledges funding from the Fritz Thyssen Foundation for supporting a research visit at the University of Pennsylvania in Spring 2020, during which part of this chapter was written.

NOTE

1. For the Google Trends service, visit https://trends.google.com/trends/. Owing to low usage, the Google Correlate service shut down on 15 December 2019.

REFERENCES


Zagheni, E. and I. Weber (2012), ‘You are where you e-mail: Using e-mail data to estimate international migration rates’, Proceedings of the 4th Annual ACM Web Science Conference, June, 348–51.
