Article

Sociological Methods & Research
2022, Vol. 51(2) 866–886

DOI: 10.1177/0049124119882480
journals.sagepub.com/home/smr


Acceptance of the Automated Online Collection of Geographical Information

Barbara Felderer[1,2] and Annelies G. Blom[3]

Abstract

The ease at which online paradata can be captured in web surveys seems to increase social researchers' desire to collect such data. Yet little attention is paid to whether respondents actually approve of their collection. This article, therefore, studies online survey respondents' acceptance of automatically collecting their geographical locations. In wave 4 of the German Internet Panel, we asked respondents for their consent to automatically track their location using a JavaScript. Respondents were also asked to report their location in a set of traditional survey questions. About 62 percent of respondents consented to the automated collection of their location whereas 97 percent provided their location manually. With respect to consent biases, we find evidence that the composition of the achieved sample of geo-located respondents is biased and that the personal characteristics associated with respondents' willingness to be geo-located differ between the automated tracking and manual provision of geo-information.

[1] GESIS - Leibniz Institute for the Social Sciences
[2] Collaborative Research Center 884 "Political Economy of Reforms," University of Mannheim, Germany
[3] Faculty of Social Sciences, University of Mannheim, Germany

Corresponding Author:
Barbara Felderer, University of Mannheim, A5, 6, Mannheim, Baden-Württemberg 68131, Germany.
Email: felderer@uni-mannheim.de

Keywords

paradata, informed consent, consent bias, geographical information, data confidentiality

Recent years have seen gradual but far-reaching changes in the survey research landscape that were initiated by some key developments: Decreasing response rates and rising survey costs (Couper 2013; Groves 2011; Hox and De Leeuw 1994) paired with increasing access to alternative data sources for the social sciences, such as Big Data from websites and social media, administrative records, and geo-coded data (Callegaro 2013; Couper 2013; Kreuter 2015; Kreuter, Müller, and Trappmann 2010). The data from these new sources are attractive because they typically come at a lower cost per unit than survey data (Groves 2011) and are available as large data sets, enabling complex statistical analyses. Such data sources can be of interest in their own right (e.g., to forecast election results; see Gayo-Avello 2013) or as an augmentation of survey interview data (Couper 2013; Couper and Singer 2013; Groves 2011), where the survey data are linked to the data set from the alternative data source via a common link-ID (for studies consent to administrative record linkage, see Korbmacher and Schröder 2013; Kreuter, Sakshaug, and Tourangeau 2015; Sakshaug and Huber 2015; Sakshaug and Kreuter 2014; Sakshaug, Tutz and Kreuter 2013; Sakshaug, Wolter, and Kreuter 2015; Sala, Knies, and Burton 2014; for a discussion of the benefits and challenges of data linkage, see Blom and Korbmacher 2018).

Paradata, which describe the survey data collection process and are collected as a by-product thereof, often accompany alternative data sources. They are particularly common in web surveys, where we can, for example, collect time stamps, clicking patterns, information about the device used, and Internet protocol (IP) addresses (Callegaro 2013) at little additional cost. IP addresses are a special type of paradata because they contain information about the survey process (e.g., variation in IP addresses across panel waves may indicate respondent mobility) and, in addition, can function as a link-ID through which geo-coded data can be linked to the survey data set.

Linking geo-coded data to the location of where respondents fill in the survey is valuable for both substantive and methodological research because such data enrich the survey data set with additional explanatory variables. Examples of these are weather or climate data and distances from public places like supermarkets, green spaces, or schools.

Some studies link weather data to survey data via a self-reported or address-based geographical link-ID and find that the weather affects survey responses. Feddersen, Metcalfe, and Wooden (2016) merged data from the Household, Income and Labour Dynamics in Australia survey to data from the Australian Bureau of Meteorology via respondents' addresses and found that the weather and climate impacted on reported life satisfaction. In a similar vein, Egan and Mullin (2012) and Shao (2017) find an effect of the local weather on survey respondents' perception of global warming.

While their research is of great value, there are two methodological challenges to these three studies: They rely on respondents' willingness and their ability to relate their location accurately by means of a manually reported address or zip code. Alternatively, researchers may automatically locate respondents during the interview. Such an automated collection of geographical locations has two advantages: First, it reduces the space needed in a questionnaire and, in consequence, the response burden. Second, it enables capturing locations for people filling in the questionnaire in places for which they do not know the address.

In surveys that use GPS-enabled devices, such as smartphones, the automated location process can take place via satellites that provide exact coordinates of the respondents' whereabouts. Such technology is now used for mobile web surveys that focus on surveying daily mobility patterns (Lin and Hsu 2014). If researchers wish to collect location data for survey respondents who participate via desktop or laptop computers, however, GPS tracking is not possible. Moreover, the vast majority of panelists in many online panels still fill in their questionnaires on desktop and laptop computers. For example, during the German Internet Panel (GIP) survey in March 2013, when we conducted our study, only 3 percent and 4 percent of the respondents filled in the questionnaire using a smartphone or a tablet, respectively. Over time, these figures have increased, however, even in January 2018 only 16 percent and 13 percent of the GIP panelists responded via smartphone or tablet, respectively. For a comprehensive picture of the panelists' location, GPS tracking, therefore, is not yet viable. Instead, collecting IP addresses remains necessary to automatically track panelists' location.

Enthusiasts of the automated location of respondents, nevertheless, tend to overlook the ethical rules of conduct and data protection regulations to which the collection, storage, and analytical use of such data are subjected. For example, the collection of respondents' IP addresses, which are device-type paradata, routinely takes place at the beginning of web interviews (Callegaro 2013), and respondents are usually not able to prevent their collection (ESOMAR 2011). However, these data are, arguably, highly personal and, thus,

underlie the same regulations as other personal identifiers like addresses or social security numbers (Callegaro 2013; ESOMAR 2011). Their collection, usage, and storage, thus, require the informed consent by the survey respondents (Couper and Singer 2013; Singer and Couper 2011).

Informed consent for the collection of paradata starts a difficult debate for survey researchers. While researchers are used to asking respondents for consent to the collection, storage, and analysis of the answers given during an interview, paradata are less tangible to respondents and, consequently, more difficult to "inform" about (Couper and Singer 2013; Singer and Couper 2011). Moreover, researchers' endeavors to use new data sources typically do not stop at the collection of paradata, particularly where IP addresses are concerned. Instead, researchers typically aspire to using the obtained IP address as a link-ID to link additional data from other sources, for example, the geo-coded weather information discussed above (Feddersen et al. 2016). Again, informed consent from the survey respondent to allow this linking process appears to be necessary.

In summary, to link geo-coded information to survey data, informed consent is needed for both the collection of a geo-link-ID and the linking process. While many web surveys routinely collect a geo-link-ID in the form of the IP address of the respondents' device, respondents are seldom informed about this and even more seldom explicitly consent to it.

Literature

Given the scarcity in published research on consent to the collection and linking of geographical information, two related strands of literature inform our research: research on informed consent to paradata collection and research on the consent to data linkage.

Even research on informed consent to paradata collection in survey interviews is still surprisingly scarce and limited to a single study by Couper and Singer (2013). In vignette experiments, they study respondents' hypothetical willingness to participate in surveys in which paradata—characteristics of the browser, key strokes, and time stamps—are collected. Respondents who agree to participate are further asked whether they would permit the use of the paradata. Varying the amount of information on paradata and their use, Couper and Singer (2013) find that any mention of paradata reduces the respondents' willingness to participate in a survey. Asking respondents for consent to use this kind of paradata in an actual survey yields consent rates between 66 percent and 72 percent, depending on the description of the paradata provided.

The literature on consent to linking individual survey records to other data sources, such as administrative or health data, is a little less scarce. In this context, several different approaches to maximize consent to data linkage have been experimentally tested.

A first set of research projects investigates the effect on consent of mentioning to respondents that allowing data linkage will reduce the number of questions needed to be asked during the survey interview. Asking for consent to link web survey responses to administrative records of the German Federal Employment Agency, Sakshaug and Kreuter (2014) find that such time-saving and interview-shortening arguments benefit consent. However, in a telephone study, Sakshaug et al. (2013) did not find an effect on linkage consent, when the consent request was motivated in terms of time savings for the respondent.

Research is also mixed when it comes to the effect on consent of loss framing, where respondents are informed that their survey responses will be less useful if no consent to data linkage is provided, versus gain framing, where respondents are informed that their survey responses will be more useful if consent to data linkage is provided. Kreuter et al. (2015), for example, find loss framing to be more effective in achieving consent than gain framing. Yet, Sakshaug et al. (2015) conclude that the effect of gain versus loss framing depends on whether the gains and losses are related to the usefulness of the information that has already been provided, or is yet to be provided, by the respondent.

Concerning the placement of the consent question, Sakshaug et al. (2013) find the consent rate to be higher when consent is requested at the beginning instead of the end of an interview. Finally, investigating correlates of consent, Korbmacher and Schröder (2013) show that consent to the collection of blood spots (biomarkers) depends on respondents' sociodemographics as well as the characteristics of the interview situation and the interviewer.

As this overview shows, surprisingly few studies have thus far investigated informed consent to the collection of online paradata despite its ubiquity and increasing importance for survey research. In particular, we know next to nothing about respondents' consent to the collection of their geolocation through automated processes.

Our article aims to fill this research gap by shedding light on respondents' consent to the automated and manual collection of geo-link-IDs during an online survey of the GIP. More specifically, we ask respondents for consent to run a JavaScript program that records their IP address and to link their geographical location to their survey data via this IP address. In addition, we ask a series of questions about the respondents' current location to detect

whether respondents resist revealing geographical information altogether or whether they only resist the automated collection thereof.

Data

As a probability-based online panel that includes previously off-line persons in order to draw inference to the general adult population in Germany, the GIP is well-suited to the study of the mechanisms of informed consent to the collection and linking of geographical information in the general population. Set up in 2012, GIP panelists were recruited in two stages. During the first stage, a strict area probability sample with prior listing and in-office sampling of household addresses was interviewed in a short face-to-face recruitment interview (AAPOR (2016) response rate (RR2): 52.1 percent). Subsequently, all household members aged 16–75 were invited to participate in the GIP online panel (cumulative response rate at panel registration: 18.5 percent).[1] To become GIP panel members, all respondents needed to give permission to the collection and storage of their survey and paradata at the beginning of their first online interview. (For information on the design and fieldwork of the GIP, see Blom, Gathmann, and Krieger 2015; Blom et al. 2016. Please note that in 2014 and 2018 additional samples were recruited into the GIP, which, however, were not included in this study.)

If, during the face-to-face recruitment interview, a household was found to lack a computer and/or Internet access, these so-called off-liners were equipped with a user-friendly computer and/or Internet connection to enable their participation in the online panel and, thereby, minimize noncoverage (for information on the representativeness of the GIP data, in particular with respect to the offline population, see Blom and Herzing 2016; Blom et al. 2017; Herzing and Blom 2018).

GIP panelists are interviewed in bimonthly online surveys of approximately 20 minutes on a variety of social, economic, and political topics. Our study was conducted at the end of wave 4, in March 2013.[2] With a completion rate of 69.7 percent, 1,118 of the 1,603 GIP panelists participated in this wave, which is equivalent to a cumulative response rate of 12.9 percent.

Thirty-four respondents broke off the questionnaire before our questions were asked (break-off rate: 3.0 percent). Five respondents skipped all location questions. These item nonrespondents were excluded leaving 1,079 respondents for the descriptive and bivariate analyses. Due to a small amount of item nonresponse in the independent variables, the sample size for the multivariate models was further reduced by 11 cases to 1,068.[3]

Figure 1. Screenshot and English translation of manual location report.

Toward the end of the questionnaire, respondents were asked to manually report the address at which they were filling in the questionnaire (city, postal code, and German state, see Figure 1).[4] Subsequently, respondents were asked for consent for a software to automatically record their location using a JavaScript plugin[5] (see Figure 2).

Analytical Strategy

With our research, we aim to contribute to the literature by addressing three research questions:

1. What is the acceptance among the general population in Germany of (a) the manual and (b) the automated collection of their geographical location?
2. Who consents to being located and who refuses, i.e. what are the characteristics of consenters and refusers?
3. Are there differences in the characteristics of people who manually report their location and people who consent to the automated collection of their geographical location?

The first research question looks into the main effects of our study. We answer this by analyzing the rate at which respondents manually provided a location (a city name, a zip code, or at least one of these indicators) and the consent rate for the automated collection of location information via the IP address. We further check whether the rate at which the geo-location is

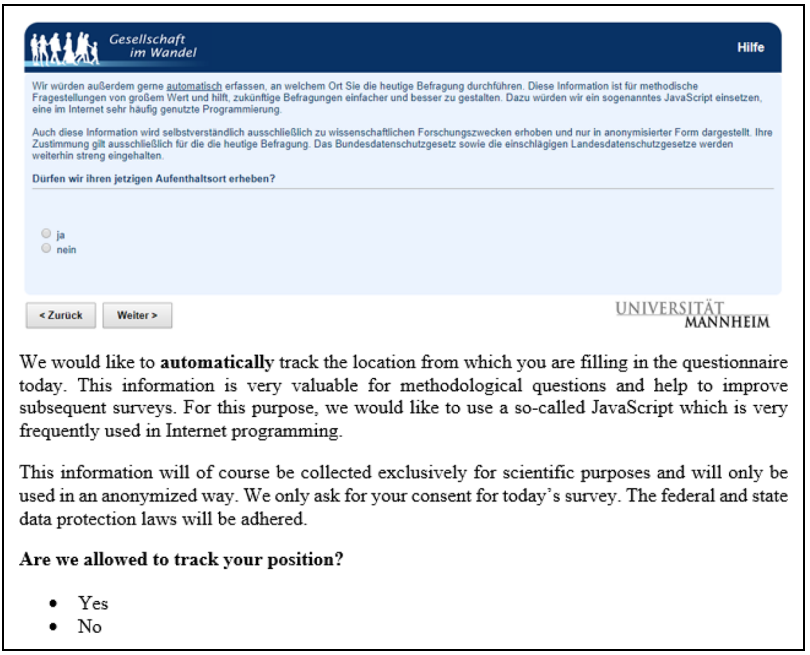

Figure 2. Screenshot and English translation of question of consent to automated location.

manually provided significantly differs from the consent rate for the automated collection of this information through paired *z*-tests.

To answer question 2, we investigate correlations between respondent characteristics and whether the location information is provided. For this purpose, we build an indicator "manual" that is 1 if respondents reported at least one of two manual location measures (city and/or zip code) and 0 if respondents did not provide either piece of information. Furthermore, we build an indicator "automated" that is 1 if respondents consented to the automated tracking of their location and 0 otherwise. These indicators are used as dependent variables in two logistic regression estimations on respondent characteristics. Our models control for the complex sampling design of the GIP by taking the clustering of individuals within households within primary sampling units into account using Jackknife variance estimation (see Lumley et al. 2004).

To answer question 3 of whether there are differential effects of respondents' characteristics on consent to the two different geo-location questions,

we model the willingness to provide a geo-location manually and consent to the automated collection of this information simultaneously using multilevel modeling. For this purpose, we reshape the data set to long format, so that each respondent has two observation rows, one for each of our dependent variables "manual" and "automated" ($N = 2{,}158$). We call both the willingness to report a geo-location and the consent to the automated measurement "compliance" and generate an indicator "question content" that denotes whether the row refers to the question about the automated or the manual collection of the geo-location. We run a logistic regression of the compliance variable on the "question content" indicator together with the same respondent characteristics as in the single models. Most importantly, we also include interaction terms of the "question type" indicator and respondent characteristics. Because the "compliance" variable is nested within respondents, we add random intercepts to our model. The significance of the interaction terms indicates differential effects of person characteristics on automated versus manual geo-location collection. Standard errors are clustered for Primary Sampling Units (PSUs), households, and respondents.

The respondent characteristics that we consider in the logistic regressions are basic characteristics that were collected in the GIP core questionnaire in wave 1. These characteristics are selected for two reasons. First, after wave 1, the participation patterns in the GIP vary greatly from wave to wave. While 43 percent of GIP panelists participate in all waves during the first two years, the rest of the panelists miss at least one wave at some point. Many of these panelists miss one or two waves but are otherwise loyal long-term GIP members. By using predictors from wave 1 only, we minimize the scope for item nonresponse.

Second and more importantly, we deliberately choose characteristics that are of general importance to researchers using the GIP and similar probability-based online panels (see Blom et al. 2016). This means that we investigate general sociodemographic backgrounds of panelists as well as indicators that are likely to be correlated with the topic focus of the GIP (i.e., social, political, and economic research). This way, we draw a profile of consenters and nonconsenters that may generalize beyond the scope of our study in the GIP and may thus inform other researchers regarding the bias trade-offs associated with collecting location information for online respondents.

If these characteristics affect the willingness to consent to the collection of geo-locations, the subsample of respondents who consented is not representative of the GIP sample. As a typical use of geo-locations is to use them as a link to outside data sources and expand the survey data set by merging

additional information on location level, findings from the reduced data set of consenters might be biased and not generalizable to the population of interest. Most importantly, bias in the personal characteristics we study will most likely lead to biases in other key survey variables that they are correlated with like opinions and attitudes.

The independent variables in our models are gender, age, education (low, medium, high), place of residence (East, West), household size (single, two plus household members), frequency of computer use for private purposes (never/less than monthly, every month, every week, every day), and the Big Five personality traits: openness to experience, conscientiousness, extroversion, agreeableness and neuroticism (from the 10-item short version of the Big Five Inventory measured on five-point scales; see Rammstedt and John 2007; see Table 1 for an overview of all characteristics used in the models).

In addition, wave 4 collected information about the type of location respondents were at, while filling in the survey. We asked whether respondents were currently on the move (e.g., on a train); at work; with family, friends, or acquaintances; at home; in a public space (e.g., in a café or restaurant); or in another place. Responses given to "in another place" were back-coded and the categories "with family, friends, or acquaintances," "in a public place," and "on the move" collapsed into a single new category "outside the home and work." Thus, the type of location indicator has three categories: at home, at work, and outside the home and work.

Results

We first analyze our first research question on the acceptance of the collection of geographical information. In total, 62.1 percent (670 respondents) consented to being located via a JavaScript plug-in, while 37.9 percent (409 respondents) refused (see Figure 3). A meaningful city name was provided by 95.9 percent (including foreign cities) and a valid postal code by 90.4 percent of the respondents. In total, 1,045 respondents (96.9 percent) reported at least one manual location measure. The rate at which the geolocation is manually provided is significantly higher than the consent rate to the automated collection of this information (paired z-test: $p < .01$).

The willingness to report a location and consent to the automated tracking is highly correlated (see Table 2). Of the 670 respondents who consented to the automated geographical data collection, only 5 did not provide a city name or zip code. And of the 1,045 respondents who manually provided location information, 665 also consented to automated geographical data collection. However, the respondents who did not provide location

Table 1. Summary of Independent Variables Used in the Regression Models.

Variable	Coding	Summary Statistics
Female	Dichotomous: 1 = female, 0 = male	49.9% female
Age	Continuous: range 18–77 years	Mean = 46.9, *SD* = 15.2
Single household	Dichotomous: 1 = single household, 0 = 2+ persons in household	14.0% single households
East Germany	Dichotomous: 1 = East Germany, 0 = West Germany	19.1% East Germany
Education	Three categories: 1 = low education, 2 = medium education, 3 = high education	21.5% low education, 35.3% medium education, 43.1% high education
Neuroticism (Big5)	Continuous: range 1–5	Mean = 2.8, *SD* = 0.9
Agreeableness (Big5)	Continuous: range 1–5	Mean = 3.0, *SD* = 0.8
Extroversion (Big5)	Continuous: range 1–5	Mean = 3.2, *SD* = 0.9
Conscientiousness (Big5)	Continuous: range 1–5	Mean = 4.0, *SD* = 0.8
Openness (Big5)	Continuous: range 1–5	Mean = 3.4, *SD* = 0.9
Location	Three categories: 1 = at home, 2 = at work, 3 = outside home and work	88.3% at home, 5.1% at work, 6.6% outside home and work
Private computer usage	Four categories: 1 = every day, 2 = every week, 3 = every month, 4 = less than monthly	62.8% every day, 24.3% every week, 7.3% every month, 5.7% less than monthly

information might systematically differ from those who did, in particular for the automated geographical data collection, where there is more variation to explain than for the manual collection.

Therefore, to answer research question 2, we run two logistic regressions of the indicators "manual" and "automated" on the respondents' characteristics, as described above. The first two columns in Table 3 show the respective results.

Although 96.9 percent of respondents report a city name or postal code, we find selectivity in the willingness to report locations. The propensity to provide geographical data manually is significantly affected by a respondent's level of extroversion: Higher levels of extroversion are associated

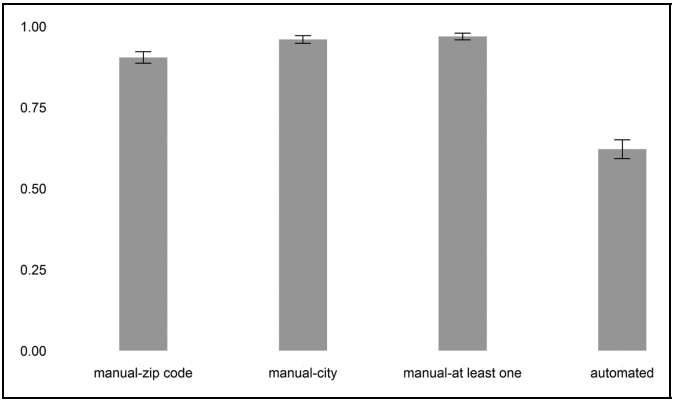

Figure 3. Consent rates for automated and manual geo-location questions (bars) with 95 percent confidence intervals (lines). $N = 1{,}079$.

Table 2. Cross-table of Willingness to Manually Provide a Geo-location and Consent to the Automated Collection of Geo-location.

	Automated Collection of Geo-location		
	No Consent	Consent	Total
Manual collection of geo-location			
Not provided	29	5	34
	2.7%	0.5%	3.2%
Provided	380	665	1,045
	35.2%	62.1%	96.9%
Total	409	670	1079
	37.9%	62.1%	100.0%

Note: Absolute numbers and cell percentages.

with higher willingness to report a city name or zip code ($p < .05$). We also find a marginally significant negative association between respondents who were out of the home and work when they took the survey and their willingness to manually provide geographical data ($p < .1$). This makes sense because these respondents have a higher chance of not knowing the address that they are at. But this is unfortunate because while respondents' home addresses are usually known to the survey organization, their location when filling in the questionnaire is not known when respondents are on the go. There is no effect of the other personality traits or sociodemographic characteristics on the propensity to manually report location information.

Table 3. Results of Logistic Regressions of Consent to Automated Geo-location and Willingness to Manually Provide a Geo-location.

	Automated		Manual		Significant Interactions With Question Type in Compliance Model
	Coefficient	Standard Error	Coefficient	Standard Error	
Female	−0.123	(.135)	−0.387	(0.398)	
Age	0.028***	(.005)	−0.010	(0.015)	**
Single household	−0.276	(.215)	1.976	(13.806)	**
East Germany	0.291	(.180)	0.900	(0.672)	
Medium education	−0.294	(.206)	−0.699	(0.717)	*ns*
High education	−0.420**	(.199)	−0.276	(0.750)	*ns*
Neuroticism	0.009	(.079)	0.087	(0.196)	
Agreeableness	0.201**	(.082)	0.296	(0.243)	*ns*
Extroversion	0.013	(.072)	0.476**	(0.194)	***
Conscientiousness	−0.049	(.102)	0.299	(0.236)	
Openness	0.152**	(.075)	−0.310	(0.227)	**
Location: at work	−0.380	(.266)	−0.885	(0.665)	*ns*
Location: outside home and work	−0.365	(.296)	−1 208*	(0.648)	*ns*
Private computer use: every day	−0.341	(.346)	−0.396	(12.136)	
Private computer use: every week	0.042	(.355)	−1.014	(12.140)	
Private computer use: every month	0.346	(.375)	0.382	(18.376)	
Constant	−1.636**	(.681)	2.567	(12.448)	
N	1,068		1,068		
McFadden's pseudo R^2	.05		.09		

Note: Models control for the complex sample design using Jackknife variance estimation.
*p < .1. **p < .05. ***p < .01.

We find that consent to the automated collection of geographical data is significantly correlated with age, education, and two of the Big Five personality traits ("openness to experience" and "agreeableness"). Consent to automated collection increases with age ($p < .01$) but is lower the higher a respondent's educational level is (difference between low and high education significant at $p < .05$). This is surprising and we do not have a clear

interpretation of this finding. One could speculate that younger and more educated respondents might be more aware of the negative sides of new technologies and thus refuse at a higher rate, but more support is needed for this. Higher levels of "openness to experience" and "agreeableness" increase the propensity to consent (both $p < .05$). The positive effect of openness to experience on automated location collection makes sense because the technology that we used was rather innovative at the time and curiosity might increase the willingness to try out new technology. Agreeableness is found to have a positive effect on consent as well. This also makes sense because more agreeable persons are less likely to refuse a researcher's request. There is no significant effect of the other Big Five personality traits, gender, household size, location details, or computer use.

Looking into research question 3 and thus testing whether the respondents' characteristics affect their willingness to report a geo-location manually and their willingness to consent to the automated geo-location collection differentially, we run a single logistic regression of the compliance indicator on the respondent characteristics including interactions with the automated versus manual location request indicator. We only estimate interaction effects with this indicator for the variables that were found to be significant in one of the two separate models or whose coefficients show opposite signs when comparing the two models. Consequently, the model included interactions with age, education, location when filling in the questionnaire, living in a single household, and the Big Five personality traits extroversion, openness, and agreeableness.

The third column of Table 3 summarizes the results of the compliance model. (The regression coefficients and standard errors are found in Table A1 in Online Appendix.) Our results show that age, living in a single household, and the Big Five personality traits extroversion and openness significantly differently affect respondents' willingness to manually report a geo-location and their consent for the automated collection of a geo-location, while we do not find significant interactions for education, respondents' locations when filling in the questionnaire, and agreeableness.

Discussion

In web surveys, we can automatically collect paradata about the survey process such as keystrokes, response times, or IP addresses. However, the ease at which online paradata can be captured seems to exceed respondents' understanding of how such data may be used by researchers (see also Couper

and Singer 2013; Singer and Couper 2011). This poses a problem for ethical requirements regarding informed consent by the survey respondent, especially when informing respondents is difficult because of the complexity of the data collection processes involved and the multitude of potential future uses of the data collected.

IP addresses, for example, are typically not of interest in their own right but serve as a link to geo-coded data sets. The resulting combined data sets enrich the survey data and increase its analytical potential. While linking geo-coded data may also be achieved by asking respondents to manually report their location, the collection of IP addresses offers a more precise location and reduces the response burden. However, respondents may perceive the automated collection of their location as intrusive and may thus object to it.

For researchers, there are obvious benefits to automatically collecting IP addresses and linking geographical information to a survey data set. However, whether respondents are fully informed about the processes involved and the uses of their data, when they agree to typical data protection statements upon joining an online panel, is less clear. This article aims to reduce this gap in the current literature.

In an online panel sample that upon registration agreed to the general collection of paradata, we analyze respondents' consent to the automated collection of their geographical location via a JavaScript plug-in that tracked their IP address, when asked specifically for consent to this procedure. In addition, respondents were asked to manually report their location in a set of standard survey questions. We compare consent to the automated location tracking to respondents' willingness to manually report their location.

Our results show that 97 percent of the respondents are willing to manually report a city or a postal code, while only 62 percent consent to the automated location tracking. Consent to the automated tracking and manual reporting were highly correlated; only five respondents (<1 percent) who consented to the automated procedure did not provide location information manually in the survey questions. Respondents' characteristics are correlated with consent to the automated collection of their geographical information. Consenters are older, lower educated, more open, and more agreeable. Furthermore, despite the low variation in respondents' willingness to manually report a city or zip code, we find personal characteristics that significantly predict the manual reporting. Respondents who manually report their location are more likely to be extrovert and less likely to be out of the home and work at the time they filled in the questionnaire. Finally, we investigate whether there are significant differences between consenters to the automated and consenters to the manual collection of location information. We

find that the effects of age, living in a single household, and the personality traits extroversion and openness are significantly different for respondents who provide location information manually and those who consent to the automated data collection.

Our research is relevant in several ways. First, it demonstrates that panelists, who give permission to the automated collection of paradata in general when they register for the online panel, may react very differently when they are informed about the collection of a specific type of paradata and its purpose at the time that the data are actually collected. Although all of the respondents to our survey had earlier given permission to automatically collect paradata, less than a third consented to the automated location tracking just nine months later.

Second, our study shows that respondents perceive the automated and the manual collection of location data very differently. While almost none of the respondents objected to providing their location manually, more than a third refused the automated location procedure. In terms of item nonresponse, this means that researchers will end up with considerably more complete data sets, if they link external data via the manual geo-link. Furthermore, the subset of respondents who provide both types of location data differs significantly from respondents who only manually report their location but refuse the automated collection. In terms of biases, this means that researchers will end up with rather different data sets, when they link geo-coded information via an automated versus a manual link.

Some caution is advised regarding generalizability to the general population regarding the size of our main effects. Although the GIP is based on a probability sample, we cannot rule out initial nonresponse and wave-on-wave attrition bias. However, any findings regarding the differential effects of manual versus automated geo-location collection (the interaction effects) are unlikely to be affected by such biases, given the quasi-experimental design of our study, in which all respondents were asked both sets of geo-location questions. It is likely that the GIP panelists are on average more cooperative than nonrespondents to the panel and panel drop-outs. This might result in an overestimation of consent in both geo-location questions, but the interaction effect is unlikely to be affected.

While our study was able to shed light on several issues, open questions still remain. For example, why do so many respondents report a city or zip code but are not willing to consent to the automated location? How are our consent questions and the technological procedures understood by respondents? And, what do respondents believe that we do with the information that

is automatically collected? Our study can only speculate about the answers to these questions.

On a technical note, our study uncovered two important caveats regarding the feasibility of an automated location tracking. First, as we discovered after the survey, IP addresses were actually only collected for 58 percent of the respondents who consented to the automated geo-location collection. For respondents who had consented, the JavaScript application opened a pop-up window on the computer screen that asked them to agree to run the JavaScript. If respondents did not click on "agree" in this pop-up window, their IP address was not collected. Unfortunately, our information on this process is very limited. While for some respondents this pop-up window may have been blocked, others simply may not have noticed it, and again others may have reconsidered their consent once they were confronted with the pop-up window.

A second technological challenge to collecting geographical information is the conversion of IP addresses into longitudes and latitudes. In our study, only 27 percent of the IP addresses were actually converted to longitudes and latitudes by the JavaScript program. While programming may have advanced since we implemented our study and may thus overcome both of these challenges to some extent, for many cases, these challenges will remain. For example, a single IP address can still represent a group of different users via VPNs one can appear in a different location from one's true location, and geo-blockers can block the transmission of the IP address altogether. The automated collection of geographical information via GPS may seem a solution, yet this is met with new technological challenges. And, in a survey setting like the GIP, where still more than 70 percent of panelists complete their surveys on laptop or desktop computers, it remains unfeasible to comprehensively record the geo-location of respondents via GPS.

To conclude, in times where the possibilities for the automated collection of online paradata seem limitless, our study aims to encourage survey researchers to reflect on respondents' acceptance of the collection of such information. We hope to have made a valuable contribution to the surprisingly sparse literature given its importance in the technological age. The many caveats and open questions that remain are indicative of a need for considerably more research that should continuously be updated as technological possibilities advance and new ethical challenges arise.

Acknowledgments

The authors gratefully acknowledge support from the Collaborative Research Center (SFB) 884 "Political Economy of Reforms" (project Z1), funded by the German Research Foundation (DFG).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study received financial support from the Collaborative Research Center (SFB) 884 "Political Economy of Reforms" (project Z1), funded by the German Research Foundation (DFG).

ORCID iD

Barbara Felderer https://orcid.org/0000-0002-1717-0415
Annelies G. Blom https://orcid.org/0000-0003-0377-301X

Supplemental Material

Supplemental material for this article is available online.

Notes

1. All German Internet Panel (GIP) response rates were calculated following AAPOR guidelines and can be retrieved from (http://reforms.uni-mannheim.de/internet_panel/Response%20rates/).
2. This article uses data from GIP wave 1 (doi:10.4232/1.12107) and wave 4 (doi:10.4232/1.12610). The GIP data are published as Scientific Use Files in the GESIS Data Archive for the Social Sciences (GESIS-DAS). They can be retrieved from (https://dbk.gesis.org/dbksearch/GDesc2.asp?no=0109&tab=&ll=10&notabs=1&db=E). However, this article researches informed consent regarding personal and sensitive data, in particular IP addresses and manually provided geographical information. For anonymity and data protection reasons, the personal and sensitive data cannot be stored at the GESIS-DAS. Instead, they can be accessed at the secure Onsite Data Access facilities of the Collaborative Research Center "Political Economy of Reforms" of the University of Mannheim, Germany.
3. In a sensitivity analysis, we kept the respondents with missing responses and ran the logistic models on the full data set after multiply imputing the missing values using chained equations (Azur et al. 2011) but did not find any substantial differences in the results.
4. Approval for the consent question was granted by the legal team at the field agency (LINK Institute). Although the study was conducted prior to the introduction of the EU General Data Protection Regulations (GDPR) in May 2018, the wording of the consent question is in line with the GDPR.

5. The JavaScript plug-in used (geoPlugin) can be found at http://www.geoplugin.com/webservices/javascript.

References

Azur, Melissa J., Elizabeth A. Stuart, Constantine Frangakis, and Philip J. Leaf. 2011. "Multiple Imputation by Chained Equations: What Is It and How Does It Work?" *International Journal of Methods in Psychiatric Research* 20(1):40-49. doi:10.1002/mpr.329

Blom, Annelies G., Michael Bosnjak, Anne Cornilleau, Anne-Sophie Cousteaux, Marcel Das, Salima Douhou, and Ulrich Krieger. 2016. "A Comparison of Four Probability-based Online and Mixed-mode Panels in Europe." *Social Science Computer Review* 34(1):8-25. doi:10.1177/0894439315574825

Blom, Annelies G., Christina Gathmann, and Ulrich Krieger. 2015. "Setting Up an Online Panel Representative of the General Population: The German Internet Panel." *Field Methods* 27:391-408. doi:10.1177/1525822X15574494

Blom, Annelies G. and Jessica M. Herzing. 2016. "Repräsentativität in einem sequentiellen mixed-mode Design." Pp. 99-118 in *Mixed-mode Befragungen*, edited by Stefanie Eifler and Frank Faulbaum. Wiesbaden, Germany: Springer VS.

Blom, Annelies G., Jessica M. Herzing, Carina Cornesse, Joseph W. Sakshaug, Ulrich Krieger, and Dayana Bossert. 2017. "Does the Recruitment of Offline Households Increase the Sample Representativeness of Probability-based Online Panels? Evidence from the German Internet Panel." *Social Science Computer Review* 35:1-23. doi:10.1177/0894439316651584

Blom, Annelies G. and Julie M. Korbmacher. 2018. "Challenges to Record Linkage in Europe." Pp. 267-74 in *The Palgrave Handbook of Survey Methodology*, edited by Vannette, David L. and Jon A. Krosnick. Cham, Switzerland: Palgrave Macmillan.

Callegaro, Mario. 2013. "Paradata in Web Surveys." Pp. 259-79 in *Improving Surveys with Paradata: Analytic Uses of Process Information*, edited by Frauke Kreuter. Hoboken, NJ: John Wiley. doi:10.1002/9781118596869.ch11

Couper, Mick P. 2013. "Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys." *Survey Research Methods* 7:145-56. doi:10.18148/srm/2013.v7i3.5751

Couper, Mick P. and Eleanor Singer. 2013. "Informed Consent for Web Paradata Use." *Survey Research Methods* 7:57-67. doi:10.18148/srm/2013.v7i1.5138

Egan, Patrick J. and Megan Mullin. 2012. "Turning Personal Experience into Political Attitudes: The Effect of Local Weather on Americans' Perceptions about Global Warming." *The Journal of Politics* 74(3):796-809. doi:10.1017/S0022381612000448

ESOMAR. 2011. "ESOMAR Guideline for Online Research." Retrieved October 16, 2019 (https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR_Guideline-for-online-research.pdf).

Feddersen, John, Robert Metcalfe, and Mark Wooden. 2016. "Subjective Wellbeing: Why Weather Matters." *Journal of the Royal Statistical Society: Series A* 179: 203-28. doi:10.1111/rssa.12118

Gayo-Avello, Daniel. 2013. "A Meta-analysis of State-of-the-art Electoral Prediction from Twitter Data." *Social Science Computer Review* 31:649-79. doi:10.1177/0894439313493979

Groves, Robert M. 2011. "Three Eras of Survey Research." *Public Opinion Quarterly* 75: 861-71. doi:10.1093/poq/nfr057

Herzing, Jesscia M. and Annelies G. Blom. 2018. "The Influence of a Person's IT Literacy on Unit Nonresponse and Attrition in an Online Panel." *Social Science Computer Review*. Published Online First on 20th May 2018. doi:10.1177/0894439318774758

Hox, Joop J. and Edith D. De Leeuw. 1994. "A Comparison of Nonresponse in Mail, Telephone, and Face-to-face Surveys." *Quality and Quantity* 28:329-44. doi:10.1007/BF01097014

Korbmacher, Julie M. and Mathis Schröder. 2013. "Consent when Linking Survey Data with Administrative Records: The Role of the Interviewer." *Survey Research Methods* 7:115-31. doi:10.18148/srm/2013.v7i2.5067

Kreuter, Frauke. 2015. *Data Collection and Inference: Opportunities and Challenges with Administrative Data and Non-probability Sources*. Reykjavik, Iceland: Keynote Speech, European Survey Research Association.

Kreuter, Frauke, Gerrit Müller, and Mark Trappmann. 2010. "Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data." *Public Opinion Quarterly* 74:880-906. doi:10.1093/poq/nfq060

Kreuter, Frauke, Joseph W. Sakshaug, and Roger Tourangeau. 2015. "The Framing of the Record linkage consent question." *International Journal of Public Opinion Researc*, 28:142-52. doi:10.1093/ijpor/edv006

Lin, Miao and Wen-Jing Hsu. 2014. "Mining GPS Data for Mobility Patterns: A Survey." *Pervasive and Mobile Computing* 12:1-16. doi:10.1016/j.pmcj.2013.06.005

Lumley, Thomas. 2004. "Analysis of Complex Survey Samples." *Journal of Statistical Software* 9:1-19. doi:10.18637/jss.v009.i08

Rammstedt, Beatrice and Oliver P. John. 2007. "Measuring Personality in one Minute or Less: A 10-item Short Version of the Big Five Inventory in English and German." *Journal of Research in Personality* 41:203-12. doi:10.1016/j.jrp.2006.02.001

Sakshaug, Joseph W. and Martina Huber. 2015. "An Evaluation of Panel Nonresponse and Linkage Consent Bias in a Survey of Employees in Germany." *Journal of Survey Statistics and Methodology* 4:71-93. doi:10.1093/jssam/smv034

Sakshaug, Joseph W. and Frauke Kreuter. 2014. "The Effect of Benefit Wording on Consent to Link Survey and Administrative Records in a Web Survey." *Public Opinion Quarterly* 78:166-76. doi:10.1093/poq/nfu001

Sakshaug, Joseph W., Valerie Tutz, and Frauke Kreuter. 2013. "Placement, Wording, and Interviewers: Identifying Correlates of Consent to Link Survey and Administrative Data." *Survey Research Methods* 7:133-144. doi:10.18148/srm/2013.v7i2.5395

Sakshaug, Joseph W., Stefanie Wolter, and Frauke Kreuter. 2015. "Obtaining Record Linkage Consent: Results from a Wording Experiment in Germany." *Survey Methods: Insights from the Field*. Retrieved October 16, 2019 (http://surveyinsights.org/?p=7288).

Sala, Emanuela, Gundi Knies, and Jonathan Burton. 2014. "Propensity to Consent to Data Linkage: Experimental Evidence on the Role of Three Survey Design Features in a UK Longitudinal Panel." *International Journal of Social Research Methodology* 17:455-73. doi:10.1080/13645579.2014.899101

Shao, Wanyun. 2017. "Weather, Climate, Politics, or God? Determinants of American Public Opinions toward Global Warming." *Environmental Politics* 26(1): 71-96. doi:10.1080/09644016.2016.1223190

Singer, Eleanor and Mick P. Couper. 2011. "Ethical Considerations in Web Surveys." Pp. 133-62 in *Social Research and the Internet*, edited by Marcel Das, Peter Ester, and Lars Kaczmirek. Boca Raton, FL: Taylor & Francis.

The American Association for Public Opinion Research. 2016. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th ed. AAPOR.

Author Biographies

Barbara Felderer is a postdoctoral researcher with the SFB 884 "Political Economy of Reforms" at the University of Mannheim. Her research interests include survey methods, in particular measurement possibilities of online surveys, nonresponse bias, the effect of respondent incentives, measurement errors, and mode effects. Her research has been published in *Public Opinion Quarterly*, *Field Methods*, *Journal of Official Statistics* and various edited volumes.

Annelies G. Blom is a professor at the Department of Political Science, School of Social Sciences, University of Mannheim and the SFB 884 "Political Economy of Reforms," University of Mannheim. Her research interests include representativeness and measurement quality in online surveys, methods of longitudinal research, nonresponse and attrition, and interviewer effects. Her research has, among other journals, been published in *Sociological Methods & Research*, *Public Opinion Quarterly*, *Journal of the Royal Statistical Society: Series A*, and *Social Science Computer Review*.