





# Linking survey and Facebook data: mechanisms of consent and linkage

Jette Schröder <sup>a</sup>, Sanja Kapidzic<sup>a</sup>, Henning Silber <sup>a</sup>, Florian Keusch <sup>b</sup>,  
Natalja Menold <sup>c</sup> and Christoph Beuthner<sup>a</sup>

<sup>a</sup>Department of Survey Design and Methodology, GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany; <sup>b</sup>Department of Sociology, School of Social Sciences, University of Mannheim, Mannheim, Germany; <sup>c</sup>Institute of Sociology, Technische Universität Dresden, Dresden, Germany

## ABSTRACT

Using a German online panel, we investigate respondents' propensity to consent to the linkage of publicly available Facebook data to their survey data and to enable linkage. We analyse the effects of three experimental variations on consent and linkage among Facebook users: (1) the consent question wording, (2) the position of the Facebook consent question in a series of consent questions, and (3) the incentive offered. To describe potential selection bias, we consider respondents' attitudes and socio-demographic characteristics. We found effects of consent question placement on consent and linkage and effects of higher incentives on linkage, but we found no effects of question wording. Our analysis also showed that linkage was dependent on privacy and data security concerns, trust in the data-collecting actor, respondents' attitudes toward surveys, and several sociodemographic characteristics.

## ARTICLE HISTORY

Received 14 December 2023  
Accepted 2 August 2024

## KEYWORDS


Consent; data linkage; online survey; Facebook; experiment

## 1. Introduction

Data from social media platforms such as Facebook and Twitter are increasingly being used for research on fundamental social science questions beyond social media itself (Ledford, 2020). These data potentially provide valuable information on individuals' behaviours and attitudes and are not generated or influenced by the research process (Silber et al., 2022; Sloan et al., 2020). However, a major limitation is that social media users are not representative of the general population. Thus, research findings based on social media data may not be generalizable (Hargittai, 2020). Combining survey data and social media data can address this limitation, especially in the case of probability-sampled surveys. Even if only some of the sampled survey participants are social media users, the combination of both types of data allows for an estimation of and adjustment for potential biases in the social media data. Another advantage of linking survey data with social media data is the possibility of complementing social media data with structured context variables collected in the survey, such as sociodemographic characteristics or attitudes. From the point of view of survey researchers, the linkage of social media data provides an opportunity to enrich survey data at low costs and to investigate new research questions (Sakshaug, 2020; Stier et al., 2020).

Despite the great potential of linking survey and social media data, these data types are rarely combined (Al Baghal, 2020; Stier et al., 2020), and only a small number of studies

**CONTACT** Jette Schröder  [jette.schroeder@gesis.org](mailto:jette.schroeder@gesis.org)  Department of Survey Design and Methodology, GESIS – Leibniz Institute for the Social Sciences, PO Box 122155, Mannheim 68072, Germany

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/13645579.2024.2391959>

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

have examined the opportunities and challenges of data linkage (Al Baghal et al., 2020; Hughes et al., 2021; Silber et al., 2022). Usually, researchers require respondents' consent to link additional data to survey data, either because of legal requirements or because of ethics committee requirements (J. Sakshaug, 2020). To successfully link data after obtaining consent, researchers may also need further information or assistance from respondents, such as access to their social media profiles or the download and transmission of their data. Both missing respondent consent and the failure of actual linkage despite consent are sources of error in linked-data studies. To minimize the risk of bias in data analysis, it is essential to keep the error from data linkage as small as possible.

This paper explores the potential of linking Facebook data with survey data. Respondents from a commercial online access panel in Germany were asked for consent to link information from their public Facebook profiles to their survey data. While many studies have asked for respondents' actual or hypothetical consent, they did not actually link the data (e.g. Bauer et al., 2019; Keusch et al., 2019). However, consent rates and actual linkage rates may differ greatly in the case of Facebook data if respondents do not take the necessary steps to enable linkage after consenting. In our study, we therefore actually linked the respondents' survey and Facebook data.

In this paper, we address the following two research questions: (1) How can researchers improve consent and linkage rates among all respondents who are Facebook users? To systematically test the possibilities of maximizing consent and linkage rates, we implemented three survey experiments, varying the following design elements: the amount of the incentive offered, the wording of the consent question, and the position of the Facebook consent question in a series of consent questions asked in the survey. (2) What respondent characteristics affect consent and linkage? To improve the understanding of bias in the linked data, it is important to know about the effects of sociodemographic and attitudinal determinants on the consent decision and the linkage success.

## 2. Background

To link additional data to survey data, several preconditions must be met (Silber et al., 2022). First, the data must exist for a respondent. For instance, Facebook data can only be linked if the respondent uses Facebook. Second, respondents' informed consent to data linkage is necessary. Furthermore, for some types of data, including data on respondents' public or private Facebook profiles, respondents must, in a third step, enable data linkage by undertaking a specific data-sharing process. For Facebook data, respondents must undertake one of the following processes: (a) download the data themselves and share them with the researcher – often called 'data donation' (Boeschoten et al., 2022); (b) help the researchers identify the correct account (for example, by messaging a research account from their Facebook account) so that the researchers themselves can collect the data; or (c) install a browser plugin to collect the data (Silber et al., 2022).

Respondents' data linkage consent is usually requested by means of a question within the survey questionnaire (Sakshaug, 2020; Sloan et al., 2020). Compared to standard survey questions, consent questions require additional effort from respondents. Respondents' consent decisions can be understood analogously to their decisions to participate in the survey. From a rational choice perspective, these decisions are the result of weighing costs and benefits (Esser, 1986; Silber et al., 2022; Singer, 2011). It is important to note that it is not the objective costs and benefits that are relevant but rather those that the respondent subjectively perceives. Weighing perceived costs and benefits is not always an effortful thought process, as the decision may simply be based on heuristics (Gigerenzer, 2008). The same theoretical considerations can be applied to actual linkage. After consenting, respondents are asked to provide further information or perform certain steps necessary to link the data. They might then consider the additional costs arising from these tasks and weigh the costs and benefits again.

Therefore, despite their previously given consent, some respondents might not comply with the tasks required for actual linkage.

## **2.1. Effects of survey design features on consent and linkage**

For survey researchers, the main interest is not to maximize consent but linkage, i.e. the number of respondents with successful linkage in the entire sample. Optimizing survey design features can increase the number of respondents with successful linkage, thereby reducing the potential for selection bias in the linked data of Facebook users compared to the entire Facebook user population in the sample, and reducing costs, as the targeted numbers are reached with fewer respondents.

The design features and respondent characteristics we are interested in are theoretically expected to affect consent. However, predicting their effect on the subsequent step, namely the linkage among consenters (i.e. respondents who have consented to linkage), is challenging. This subset of respondents is selective based on the factors influencing the consent decision, making it difficult to anticipate whether a factor affecting consent will also influence linkage among consenters. Therefore, we only investigate this subset exploratively without specific hypotheses. However, we see no reason to expect a difference between the hypotheses regarding consent and linkage for the entire sample of Facebook users. Therefore, a variable's positive effect on consent is expected to translate to a positive effect on linkage in the entire sample. Consequently, our hypotheses pertain to consent and linkage for all Facebook users.

### **2.1.1. Wording of the consent question**

Past research has shown that wording and framing of the consent question can influence consent rates for the linkage of different data types. Experimental research has examined whether consent rates are impacted by 'benefit framing,' emphasizing the advantages of giving consent, or 'loss framing,' emphasizing the disadvantages of not giving consent (Sakshaug, 2020). Results have been mixed. For benefit framing, some studies found positive effects (Bates et al., 2012; Sakshaug & Kreuter, 2014), whereas others found no effect (Pascale, 2011; Sakshaug et al., 2013). Comparing the two framing approaches also shows mixed results: In some experiments, loss framing resulted in higher consent rates (Kreuter et al., 2016; Sakshaug et al., 2019), whereas in others, benefit-framing performed better (Sakshaug et al., 2015), or no differences were found (Sakshaug et al., 2019).

From a rational choice perspective (Singer, 2011), emphasizing benefits may influence consent likelihood by prompting respondents to consider a positive aspect they might not have otherwise thought of when weighing costs and benefits. Thus, we expect that emphasizing a particular benefit will increase respondents' likelihood of agreeing to Facebook data linkage, as well as the likelihood of linkage.

**H1a:** Benefit framing of the consent question increases the likelihood of consent to the linkage of Facebook data.

**H1b:** Benefit framing of the consent question increases the likelihood of Facebook data linkage.

### **2.1.2. Consent question sequence**

When conducting a survey involving multiple types of data linkage, a separate consent question is required for each data type. If the placement of a specific consent question in the sequence affects the consent likelihood, survey designers can optimize the question order based on their priorities.

A vignette experiment by Keusch et al. (2019) indicated that the order of consent questions can matter: Respondents received eight vignettes describing a hypothetical study involving data collection via an app on their phones. Respondents' willingness to consent was significantly higher for the first randomly assigned vignette compared to the subsequent ones. A study by Walzenbach et al. (2022) explored the impact of question order on consent to data linkage, examining whether

starting with a high or low consent probability question made a difference. In one of the two experiments, starting with a high-probability question resulted in higher consent rates. Regarding general order effects, the analyses did not consistently show declining consent rates as questions appeared later in the series. However, due to the lack of full randomization in question order, the authors could not draw definitive conclusions.

Theoretically, each additional linkage potentially adds to the respondent burden (Bradburn, 1978). Additional linked data may help researchers interpret existing data but could raise privacy concerns among respondents for the same reason, so the cost of providing consent increases with each additional request. Therefore, we expect that the likelihood of consent and linkage will decrease as more data linkage consent questions precede the Facebook consent question.

**H2a:** The more consent questions are asked before the Facebook consent question, the lower the likelihood of consent to the linkage of Facebook data.

**H2b:** The more consent questions are asked before the Facebook consent question, the lower the likelihood of Facebook data linkage.

### **2.1.3. Incentives**

The positive effects of incentives on survey participation have been widely demonstrated (Singer & Ye, 2013; Witte et al., 2021). While the use of incentives for survey participation is quite common in many countries, including Germany (Pforr et al., 2015), we are not aware of the existence of many studies incentivizing data linkage. Some studies have, however, examined the impact of incentives on consent to data linkage. For health record linkage, Ni et al. (2017) found no higher consent likelihood when gift incentives were offered. By contrast, Stone et al. (2013) found positive effects of monetary incentives for survey participation on consent to administrative record linkage within the survey. The vignette study of Keusch et al. (2019) revealed an increased willingness to participate in hypothetical studies involving automatic mobile data collection with the offer of a €10 incentive compared to no incentive.

From a rational choice perspective, an incentive promised for linkage should increase the perceived benefits of consent as well as linkage. We therefore assume a positive effect of an incentive for linkage on both the likelihood of consent and of linkage.

**H3a:** An incentive for linkage increases the likelihood of consent to the linkage of Facebook data.

**H3b:** An incentive for linkage increases the likelihood of Facebook data linkage.

## **2.2. Effects of respondent characteristics and attitudes on consent and linkage**

Sociodemographic characteristics and linkage consent have been found to be associated in many studies, but results are inconsistent when it comes to the nature of that relationship for specific characteristics (Jäckle et al., 2021; Sakshaug, 2020; Sala et al., 2012). As there are no clear theoretical arguments for the effects of sociodemographic characteristics on consent to link Facebook data, we refrain from formulating expectations. However, because their effects are nevertheless important due to the potential selection bias of Facebook users with linkage compared to the entire population of Facebook users in the sample, we analyse sociodemographic variables in an explorative way. The same applies to respondents' usage behaviour on Facebook and on social media in general.

For several attitudes, on the other hand, clear expectations can be formulated. When asked about their decision not to consent, respondents often cited concerns about data confidentiality (Sala et al., 2014). Accordingly, Sakshaug et al. (2012), as well as Mostafa (2016), found that respondents with lower privacy concerns were more likely to consent to linkage of administrative data, and Liu

et al. (2024) found that respondents with lower privacy and security concerns were more likely to consent to Twitter data linkage. Keusch et al. (2019) found that general privacy concerns did not significantly affect the willingness to participate in passive mobile data collection. Respondents' data security concerns regarding the use of a research app were, however, controlled for in the analyses and had a negative effect. Thus, they might have mediated the effect of general privacy concerns. Privacy concerns may add to the perceived costs of consent, thereby reducing the likelihood of consent to linkage and actual linkage of Facebook data. Therefore, we hypothesize:

**H4a:** The greater Facebook users' privacy concerns are, the less likely they are to consent to Facebook data linkage.

**H4b:** The greater Facebook users' privacy concerns are, the lower the likelihood of Facebook data linkage.

Theoretically, general trust and trust in the data-collecting actor should have a positive effect on consent and data linkage because more trusting respondents may feel more confident about the beneficial use of their data and privacy protection. This may lead to higher perceived benefits and/or lower costs and therefore a greater likelihood of consent and linkage. Correspondingly, Keusch et al. (2019) found respondents' trust that the research organization and other organizations would not share their data had a positive effect on the willingness to participate in passive mobile data collection. However, Bauer et al. (2019) found no effect of respondents' trust that university researchers would not share the data on the likelihood of consent to the linkage of administrative data to the survey data. We formulate the following hypotheses in line with these theoretical considerations.

**H5a:** The greater Facebook users' trust levels are, the more likely they are to give consent to Facebook data linkage.

**H5b:** The greater Facebook users' trust levels are, the higher the likelihood of Facebook data linkage.

Perceived benefits of data linkage might be higher for respondents with more positive attitudes toward surveys in general (Loosveldt & Storms, 2008). This assumption is supported by Revilla et al. (2019) findings that a positive attitude toward surveys was positively associated with respondents' willingness to share additional types of data. Thus, we hypothesize:

**H6a:** The more positive Facebook users' attitudes toward surveys are, the more likely they are to consent to Facebook data linkage.

**H6b:** The more positive Facebook users' attitudes toward surveys are, the higher the likelihood of Facebook data linkage.

### 3. Methods

#### 3.1. Data

To test our hypotheses, we used data from a 2018 online survey. Participants were recruited from a German non-probability online access panel, with quotas for gender, age, education, and federal state mirroring the general population's demographics.<sup>1</sup> Each respondent received a €2.50 incentive in the form of access panel reward points upon survey completion.

The questionnaire included questions and experiments on consent to data linkage for Facebook and several other data types like health-related and bank data (Beuthner et al., 2023). The survey also included other questions and experiments. In one of these experiments, respondents were randomly assigned to complete the survey on a desktop/laptop computer or on a smartphone. If they failed to use the designated device for the survey, they were screened out.

The module on data linkage consent began with an introductory page explaining data protection and the consent process. Before this module, respondents were randomly assigned to three framing conditions and, independently of the framing, two incentive conditions (first part of fieldwork period) or three incentive conditions (second part of fieldwork period).<sup>2</sup> The introduction page contained the framing text and an incentive announcement (without specifying its value) for eligible respondents.

In the next step, respondents received seven consecutive questions on separate screens asking for consent to data linkage for seven data domains, including Facebook data. The question order was randomized. If respondents declined to consent to Facebook data linkage, they were informed that no linkage data would be collected. Respondents who consented to Facebook data linkage received a randomly generated four-digit number and were instructed to log into their Facebook account, visit the study's Facebook page, and send their assigned number via a private message. This allowed the researchers to locate the participants' public Facebook profiles. The experimental groups that were offered an incentive were additionally told that they would receive access panel reward points worth €1 or €3 as compensation. Appendix A, Section A1 includes details of the instrument wording.

Of the 50,063 online access panel members invited to participate, 6,750 accessed the questionnaire. Of these, 2,838 (42%) were screened out for using an unassigned device, and 538 (8%) broke off before completion. This yielded a sample of 3,374 completed interviews, with a median completion time of 30 min. The analysis included only respondents who reported using Facebook more often than 'never,' accounting for 73% of the sample (2,457 cases). Due to missing values of dependent variables ( $N = 14$ ) and independent variables ( $N = 71$ ), the final analysis sample comprised 2,372 cases.

### **3.2. Analysis**

We applied logistic regressions to analyse the effects of survey and respondent characteristics on consent and linkage. We estimated three models. Model 1 tests the hypotheses regarding consent likelihood and Model 2 regarding linkage likelihood in the entire sample of Facebook users. An additional exploratory analysis analysed linkage likelihood within the subsample of respondents who consented to linkage (Model 3).

Model 2 on linkage for all Facebook users is the most important model for practitioners interested in maximising the number of respondents with linked Facebook data. This model shows which treatments impact the final proportion of Facebook users with linked data. Model 1 (consent among Facebook users) and Model 3 (linkage among consenters) provide insight into the mechanisms by which a variable increases linkage in the sample of Facebook users: They help determine whether a variable affects overall linkage by increasing the likelihood of consent (as suggested by our hypotheses) or by increasing the likelihood of linkage among consenters.

We report average marginal effects (AME) to facilitate interpretations (Best & Wolf, 2015). While odds ratios and logits are hard to assess in terms of probability changes, AMEs can be interpreted as average percentage point changes in the probability of the outcome. All analyses were conducted using Stata 17.0.

### 3.3. Measures

As dependent variables, we used two measures: (a) *consent*, coded 1 for consent to Facebook data linkage and 0 for no consent; (b) *linkage*, coded 1 for successful linkage – the respondent sent their unique code via their Facebook account and the data could be linked – and 0 otherwise. For descriptive statistics of all dependent and independent variables, see Appendix B, Table B1; for the wording of items used, see Appendix A.

The benefit-framing experiment (H1a, H1b) varied the reasons for the requested data linkage across three groups. The reason the first group received was ‘time saving,’ the second group was informed ‘for scientific purposes,’ the control group received no specific reason for data linkage.

The question-order experiment (H2a, H2b) randomly assigned the order of the consecutive consent questions for the seven data domains. We used the variable ‘position of consent question’ to indicate the Facebook consent question’s position within the sequence.

The incentive experiment (H3a, H3b) comprised three groups: those offered *no incentive*, a €1 incentive, and a €3 incentive.

Respondents’ privacy concerns (H4a, H4b) were captured with two variables. *General privacy concerns* were measured with the question: ‘How concerned are you about your privacy in general?’ on a 4-point rating scale ranging from *very concerned* to *not concerned at all*. The scale direction was reversed for analysis. The second variable focused on data security concerns in the case of smartphone data collection; it was used as an indicator of respondents’ concerns about digital data collection. Respondents rated their level of concern for eight research-related activities, such as downloading an app that collects smartphone data, on a 4-point scale from *not concerned at all* to *very concerned*. The variable *data security concerns—smartphone data collection* is an index representing the average rating of the eight items and ranges from 1 (*low*) to 4 (*high*).

The variable *interpersonal trust* (H5a, H5b) is an index based on the average ratings of three items and ranges from 1 (*low trust*) to 5 (*high trust*) (Beierlein et al., 2014). *Trust in the data-collecting actor* was measured with a general item on trust in university researchers and rated on a 4-point scale ranging from *trust fully* to *do not trust at all*, which was reversed for analysis.

The variable *attitude toward surveys* (H6a, H6b) is an index ranging from 1 (*low*) to 7 (*high*) based on the average rating of seven items on a 7-point scale ranging from *fully agree* to *do not agree at all*. The scale direction was reversed for positively worded items before creating the index.

We used two variables to measure respondents’ social media use: (a) *number of social media platforms used*, counting how many of the 13 platforms respondents used more often than ‘never’ and (b) *daily use of Facebook*, coded 1 if respondents reported using Facebook at least daily and 0 otherwise.

We used the following sociodemographic characteristics: *gender*; *age* in years; *level of education* (low, medium, high); *net monthly income*, with the categories low (up to €1,500), medium (€1,501 to €2,500) and high (€2,501 or more); and two variables regarding respondents’ place of residence: *Eastern Germany* (1 = yes, 0 = no) and *urbanicity* (1 = city with more than 100,000 residents, 0 = city or town with 100,000 or fewer residents).

## 4. Results

Out of 2,372 Facebook users in the analysis, 822 (35%) consented to data linkage, and 447 (19% of users, 54% of consenters) completed the process necessary for the data to be linked. Table 1 shows logistic model results for consent and linkage. Regarding linkage, our hypotheses focused on linkage in the sample of all Facebook users (see Model 2). In contrast, Model 3 is an exploratory analysis of linkage in the selective subsample of consenters.

The benefit framing of the consent question did not have the expected effects. Facebook users who received either reason for data linkage did not differ in their probability of consent (Model 1)

**Table 1.** Average marginal effects (AME) of survey characteristics and respondent characteristics on consent to linkage and linkage of Facebook data to survey data.

Variable	Model 1		Model 2		Model 3	
	Consent among Facebook users		Linkage among Facebook users		Linkage among consenters	
	AME	SE	AME	SE	AME	SE
<b>Survey characteristics</b>						
Framing of the consent question (ref.: no reason)						
Scientific purpose	-0.013	(0.023)	-0.014	(0.019)	-0.019	(0.040)
Time saving	-0.032	(0.023)	-0.013	(0.019)	0.034	(0.041)
Position of consent question (1-7)	-0.046***	(0.004)	-0.021***	(0.004)	0.010	(0.008)
Incentive (ref.: no incentive)						
€1	0.011	(0.021)	0.024	(0.017)	0.063	(0.038)
€3	-0.005	(0.026)	0.054*	(0.022)	0.163***	(0.047)
<b>Attitudes</b>						
General privacy concerns (1-4)	-0.005	(0.012)	-0.022*	(0.010)	-0.046*	(0.022)
Index data security concerns, smartphone data collection (1-4)	-0.119***	(0.015)	-0.091***	(0.013)	-0.065*	(0.027)
Interpersonal trust (1-5)	-0.010	(0.012)	-0.010	(0.010)	-0.006	(0.022)
Trust in data-collecting actor (1-4)	0.040**	(0.013)	0.037***	(0.011)	0.042	(0.024)
Positive attitude toward surveys (1-7)	0.025*	(0.010)	0.029**	(0.009)	0.042*	(0.018)
<b>Use of social media</b>						
Number of social media platforms used (1-13)	0.017***	(0.004)	-0.000	(0.003)	-0.020**	(0.006)
Daily use of Facebook (0 = no, 1 = yes)	0.009	(0.020)	0.011	(0.017)	0.011	(0.037)
<b>Sociodemographic characteristics</b>						
Gender: male	0.093***	(0.019)	0.068***	(0.016)	0.052	(0.034)
Age	0.002*	(0.001)	0.001	(0.001)	-0.001	(0.001)
Level of education (ref.: low)						
Medium	-0.006	(0.027)	0.052*	(0.021)	0.137**	(0.046)
High	-0.032	(0.026)	0.033	(0.021)	0.127**	(0.047)
Income (ref.: low)						
Medium	-0.003	(0.022)	-0.005	(0.020)	0.001	(0.040)
High	-0.028	(0.024)	-0.069***	(0.019)	-0.127**	(0.043)
Urbanicity	-0.041*	(0.020)	-0.035*	(0.017)	-0.042	(0.037)
Eastern Germany	0.025	(0.021)	0.021	(0.017)	0.032	(0.038)
N	2,372		2,372		822	
McFadden's Pseudo R <sup>2</sup>	0.095		0.090		0.077	

\* $p < 0.05$ . \*\* $p < 0.01$ . \*\*\* $p < 0.001$ .

or in the probability of Facebook data linkage (Model 2) from those who were not offered any reason. Neither H1a nor H1b is supported.

Consent and linkage probability among Facebook users both declined with the number of prior consent questions. Hence, our data support H2a and H2b. Since potential non-linear effects could be crucial for survey researchers when deciding on the placement of the consent question within a series of consent questions, we conducted an additional analysis modelling the position of the consent question as non-linear with a set of dummies (see Appendix B, Table B2). The analysis revealed that the decrease in consent and linkage probability from the first to the second question was quite large, whereas later positions were associated with a smaller decrease in consent and linkage probabilities.

An incentive of €1 or €3 did not increase consent likelihood compared with no incentive (Model 1). Yet, a €3 incentive raised the probability of linkage among Facebook users by five percentage points compared to no incentive (Model 2). The effect of the €1 incentive on linkage was positive, but not significant. Hence, we found partial support for H3b, which predicted a positive effect of incentives on the likelihood of linkage among Facebook users. However, H3a, which suggested a positive impact of incentives on consent likelihood, is not supported. Model 3 informs us that this difference is due to the fact that among those who had consented to data linkage, a €3 incentive raised the probability of actual linkage by 16 percentage points compared to no incentive.

Privacy concerns affected consent and linkage. Both general privacy concerns and data security concerns about smartphone data collection had the expected negative effects on consent (Model 1) and linkage (Model 2). Of these effects, only the effect of the general privacy concerns on consent failed to reach significance. Therefore, H4a and H4b, predicting a negative effect of privacy concerns on the likelihood of consent and linkage, are supported.

Of the two trust variables, only trust in the data-collecting actor showed the expected effects on consent and linkage among Facebook users (Models 1 and 2). In contrast, interpersonal trust did not yield significant effects in both models. As the effect of interpersonal trust might be mediated by trust in the data-collecting actor, we also estimated the models without this variable. Even then, the effect of interpersonal trust failed to achieve significance. H5a and H5b are supported for trust in the data-collecting actor, but not for interpersonal trust. Furthermore, a more positive attitude toward surveys was associated with a higher likelihood of consent and linkage, thus supporting H6a and H6b.

Regarding social media usage, the frequency of Facebook use did not impact consent or linkage. While the number of social media platforms used positively influenced consent likelihood (Model 1), it did not affect linkage likelihood (Model 2). This discrepancy stems from the negative impact of the number of platforms on linkage among consenters (Model 3).

Regarding sociodemographic characteristics, we observed the following effects for consent and linkage in the sample of Facebook users: Males had a greater likelihood of consent (Model 1) and linkage (Model 2). While age positively influenced consent, it did not translate into a higher linkage likelihood. Education had no impact on consent, but on linkage. The exploratory analysis revealed that the difference is due to the fact that education had a substantial impact on linkage among consenters (Model 3).<sup>3</sup> Higher income tends to be associated with lower consent and linkage likelihood. However, only the latter effect was significant. Lastly, urbanicity had a negative effect on consent and linkage likelihood. We also ran models without controlling for attitudes and social media use, finding no additional significant effects of sociodemographics (see Appendix B, Table B3).

## 5. Discussion

The present study addressed the following two questions: (1) How can researchers increase consent to linkage and linkage of Facebook data to survey data? and (2) What respondent characteristics influence consent and linkage?

To answer the first question, we experimentally tested the effects of consent question wording, the position of the Facebook consent question in a series of consent questions, and the amount of incentive offered on consent and linkage among Facebook users. To shed light on potential differences between effects on consent and linkage likelihood, we additionally conducted exploratory analyses on linkage within the selective subsample of respondents who had already given consent.

Regarding the placement of the Facebook consent question in the series, the likelihood of consent and linkage for all Facebook users decreased with the number of prior consent questions. The decline was particularly marked between the first and second positions of the Facebook consent question, as additional analyses showed.

Offering an incentive of €1 had no effect on either consent or linkage compared with offering no incentive. However, the €3 incentive increased linkage. The difference between the effect of a €3 incentive on consent and linkage is due to the fact that in the subsample of consenters a €3 incentive had a strong positive effect on linkage, as exploratory analysis showed. These findings suggest that a sufficiently large incentive can motivate consenters to complete the linkage process. Regarding the missing effect of the incentives on consent, it should be noted that the incentive was mentioned before the Facebook consent question, but the exact amount was not disclosed until afterwards. Especially for the €3 incentive, mentioning the amount earlier might have affected consent.

The wording of the consent question – that is, whether one of the two beneficial reasons for linkage was provided – affected neither consent nor linkage likelihood. This result adds to the inconsistent findings on benefit framing (Sakshaug, 2020). In our experiment, the benefit stimulus was part of the introductory text preceding the series of consent questions. Additionally, it was introduced with only 3 or 12 words, respectively, which were added to the neutrally formulated sentence used for the control group. Successful benefit framing might require a more direct and/or stronger stimulus than the one implemented in our study.

The analysis revealed that respondent characteristics affected consent and linkage likelihood, introducing selection bias in the linked data of Facebook users compared to the entire population of Facebook users in the sample. Linkage was more likely for Facebook users with lower privacy and data security concerns, higher trust in the data-collecting actor, and a more positive attitude toward surveys. Interestingly, general privacy concerns did not significantly affect consent, but had a significant effect on linkage among consenters. Therefore, it might be beneficial to re-address data security when providing linkage instructions post-consent.

The number of social media platforms used by respondents had a positive effect on consent, but no effect on linkage. The analysis of linkage in the subsample of consenters showed that the number of social media platforms had a negative effect on linkage in the selective group of consenters, which explains this difference. We can only speculate whether the latter effect could be due to time constraints associated with activities on multiple social media platforms or another mechanism.

Linkage was selective with respect to Facebook users' sociodemographic characteristics. Successful linkage was less likely for females, respondents with a low level of education (compared with a medium level), respondents with a high income, and those living in an urban area. Lower education did not affect consent rates but reduced linkage likelihood among consenters, possibly due to the perceived complexity of the procedure.

Given the described selectivity, the resulting non-linkage bias should be considered when analysing the linked data of Facebook users. In the case of nonresponse bias in surveys, the information available to assess and correct that bias is often limited. This is different in the case of the non-linkage bias, as survey responses are available for both linkers and non-linkers, allowing the creation of correction weights. For this purpose, all factors affecting linkage should ideally be collected.

Our study has several limitations. First, we recruited participants via an online access panel, which is susceptible to self-selection bias due to its non-random sampling. Indeed, the share of Facebook users in the panel was much higher than the estimate for the general population. Additionally, panel members might have a higher willingness to link to Facebook compared to the general population. Consequently, we believe that the consent and linkage rates among Facebook users observed in our study would represent an upper limit if a similar approach were employed in a randomly sampled survey in Germany.

Our study primarily aimed to estimate the causal effects of study features on consent for sharing Facebook data and on the linkage of this data to survey data. The experimental design gives us confidence in the internal validity of our results. However, we cannot rule out the possibility that individual design features could have a different effect on a probability sample of the general population. For example, the general population may be less responsive to incentives than individuals participating in an online access panel or a relevant proportion may perceive the incentives for linkage as inappropriate, reducing the positive effect of incentives. Regarding the effect of respondent characteristics, we would expect stronger rather than weaker effects in a random sample due to its potentially lesser selectivity. In addition to the type of sample, the survey mode could potentially moderate the effect of design features and respondent characteristics. To assess the generalizability of our findings, future studies should replicate this experiment in different populations and with different modes.

It would be interesting to repeat our study with private Facebook profiles. However, this would entail respondents befriending a study-specific profile, likely resulting in even lower linkage rates.

Data donation, an approach possibly perceived as less intrusive, is an option that might be worth investigating further.

In conclusion, researchers planning the linkage of Facebook data should carefully design their studies to maximize linkage rates. If several consent questions are asked within a study, but the linkage of Facebook data is the focus, the Facebook linkage question should be posed first. Offering a sufficiently large incentive for Facebook linkage is advisable (in our study, €3 in addition to €2.50 for survey completion). Finally, it is necessary to remember that linkage is selective and dependent on attitudes and socio-demographic characteristics. Especially in surveys with probability sampling aimed at inferring all Facebook users, relevant selection factors should be gathered in the survey to enable appropriate data weighting.

## Notes

1. An ethics approval for this type of survey study is not required by national law or the researchers' institution.
2. Preliminary data analysis during the fieldwork period revealed no incentive effect on data linkage consent. Therefore, we introduced a third group with a higher incentive for later-invited respondents.
3. Without controlling for income, the direction of effects and their general significance stayed the same (not shown).

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Notes on contributors

*Jette Schröder* is a senior researcher at the Survey Design and Methodology Department at GESIS - Leibniz Institute for the Social Sciences in Mannheim, Germany. Her research interests include survey methodology, well-being, and family research.

*Sanja Kapidžić* is a researcher at GESIS – Leibniz-Institute for the Social Sciences in Mannheim, Germany. Her research interests include social media communication, self-presentation in online environments, and individual differences in online interactions.

*Henning Silber* is a Senior Researcher and head of the Survey Operations Team at the Survey Design and Methodology Department at GESIS – Leibniz Institute for the Social Sciences. His research interests include survey methodology, political sociology, and the experimental social sciences.

*Florian Keusch* is Professor of Social Data Science and Methodology at the University of Mannheim, Germany. In his research, he assesses the quality of modern methods of collecting data for the behavioral and social sciences. He is particularly interested in how to best leverage the collection of self-reports and digital behavioral data, for example, from smartphones, wearables, online and device log files, and Internet search queries, to better answer research questions on migration, employment, and inequality.

*Professor Dr. Natalja Menold* completed a Master's degree in psychology at the University of Tuebingen in 2000. She received her doctorate from the University of Dortmund in 2006 and her *Venia Legendi* from the University of Mannheim in 2017. She has been Chair of Methods in Empirical Social Research at the Institute of Sociology, Faculty of Arts, Humanities and Social Science, Dresden University of Technology since April 2019. Her research interests include among others survey methodology and measurement error.

*Christoph Beuthner* was an Associate Researcher at GESIS. He obtained his PhD from the University of Mannheim. His research interests include machine and deep learning.

## ORCID

Jette Schröder  <http://orcid.org/0000-0002-1000-5855>

Henning Silber  <http://orcid.org/0000-0002-3568-3257>

Florian Keusch  <http://orcid.org/0000-0003-1002-4092>

## References

- Al Baghal, T. (2020). Linking survey and social media data. *Understanding society working paper series, no. 2020-04*. Economic and Social Research Council. <https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2020-04.pdf>
- Al Baghal, T., Sloan, L., Jessop, C., Williams, M. L., & Burnap, P. (2020). Linking Twitter and survey data: The impact of survey mode and demographics on consent rates across three UK studies. *Social Science Computer Review*, 38(5), 517–532. <https://doi.org/10.1177/0894439319828011>
- Bates, N., Wroblewski, M. J., & Pascale, J. (2012). Public attitudes toward the use of administrative records in the U.S. Census: Does question frame matter? *Research report series – survey methodology no. 2012-04*. United States Census Bureau. <https://www.census.gov/content/dam/Census/library/working-papers/2012/adrm/rsm2012-04.pdf>
- Bauer, P. C., Keusch, F., Kreuter, F., & Capraro, V. (2019). Trust and cooperative behavior: Evidence from the realm of data-sharing. *PLOS ONE*, 14(8), e0220115. <https://doi.org/10.1371/journal.pone.0220115>
- Beierlein, C., Kemper, C., Kovaleva, A., & Rammstedt, B. (2014). Interpersonales Vertrauen (KUSIV3). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. <https://doi.org/10.6102/zis37>
- Best, H., & Wolf, C. (2015). Logistic regression. In H. Best & C. Wolf (Eds.), *The SAGE handbook of regression analysis and causal inference* (pp. 153–172). SAGE Publications.
- Beuthner, C., Weiß, B., Silber, H., Keusch, F., & Schröder, J. (2023). Consent to data linkage for different data domains – The role of question order, question wording, and incentives. *International Journal of Social Research Methodology*. Advance online publication. <https://doi.org/10.1080/13645579.2023.2173847>
- Boeschoten, L., Mendrik, A., van der Veen, E., Vloothuis, J., Hu, H., Voorvaart, R., & Oberski, D. L. (2022). Privacy-preserving local analysis of digital trace data: A proof-of-concept. *Patterns*, 3(3), 100444. <https://doi.org/10.1016/j.patter.2022.100444>
- Bradburn, N. (1978). Respondent burden. *Proceedings of the Survey Research Methods Section of the American Statistical Association* (Vol. 35, pp. 35–40). American Statistical Association. [http://www.asasrms.org/Proceedings/papers/1978\\_007.pdf](http://www.asasrms.org/Proceedings/papers/1978_007.pdf)
- Esser, H. (1986). Über die Teilnahme an Befragungen. *ZUMA Nachrichten*, 10(18), 38–47. <https://www.ssoar.info/ssoar/handle/document/21030>
- Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, 3(1), 20–29. <https://doi.org/10.1111/j.1745-6916.2008.00058.x>
- Hargittai, E. (2020). Potential biases in big data: Omitted voices on social media. *Social Science Computer Review*, 38(1), 10–24. <https://doi.org/10.1177/0894439318788322>
- Hughes, A. G., McCabe, S. D., Hobbs, W. R., Remy, E., Shah, S., & Lazer, D. M. J. (2021). Using administrative records and survey data to construct samples of tweeters and tweets. *Public Opinion Quarterly*, 85(S1), 323–346. <https://doi.org/10.1093/poq/nfab020>
- Jäckle, A., Burton, J., Couper, M. P., Crossley, T. F., & Walzenbach, S. (2021). Understanding and improving data linkage consent in surveys. *Understanding society working paper series no. 2021-01*. Economic and Social Research Council. <https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2021-01.pdf>
- Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly*, 83(S1), 210–235. <https://doi.org/10.1093/poq/nfz007>
- Kreuter, F., Sakshaug, J. W., & Tourangeau, R. (2016). The framing of the record linkage consent question. *International Journal of Public Opinion Research*, 28(1), 142–152. <https://doi.org/10.1093/ijpor/edv006>
- Ledford, H. (2020). How data from Facebook, Twitter and other data troves are revolutionizing social science. *Nature*, 582(7812), 328–330. <https://doi.org/10.1038/d41586-020-01747-1>
- Liu, S., Sloan, L., Al Baghal, T., Williams, M., Jessop, C., & Seródio, P. (2024). Linking survey with Twitter data: Examining associations among smartphone usage, privacy concern and Twitter linkage consent. *International Journal of Social Research Methodology*. Advance online publication. 1–15. <https://doi.org/10.1080/13645579.2023.2299482>
- Loosveldt, G., & Storms, V. (2008). Measuring public opinions about surveys. *International Journal of Public Opinion Research*, 20(1), 74–89. <https://doi.org/10.1093/ijpor/edn006>
- Mostafa, T. (2016). Variation within households in consent to link survey data to administrative records: Evidence from the UK millennium cohort study. *International Journal of Social Research Methodology*, 19(3), 355–375. <https://doi.org/10.1080/13645579.2015.1019264>
- Ni, M. Y., Li, T. K., Hui, R. W., McDowell, I., & Leung, G. M. (2017). Requesting a unique personal identifier or providing a souvenir incentive did not affect overall consent to health record linkage: Evidence from an RCT nested within a cohort. *Journal of Clinical Epidemiology*, 84, 142–149. <https://doi.org/10.1016/j.jclinepi.2017.01.003>

- Pascale, J. (2011). Requesting consent to link survey data to administrative records: Results from a split-ballot experiment in the survey of health insurance and program participation (SHIP). *Study series – survey methodology no. 2011-03*. United States Census Bureau. <https://www.census.gov/content/dam/Census/library/working-papers/2011/adrm/ssm2011-03.pdf>
- Pfarr, K., Blohm, M., Blom, A., Erdel, B., Felderer, B., Fräßdorf, M., Hajek, K., Helmschrott, S., Kleinert, C., Koch, A., Krieger, U., Kroh, M., Martin, S., Saßenroth, D., Schmiedeberg, C., Trüdinger, E.-M., & Rammstedt, B. (2015). Are incentive effects on response rates and nonresponse bias in large-scale, face-to-face surveys generalizable to Germany?: Evidence from ten experiments. *Public Opinion Quarterly*, 79(3), 740–768. <https://doi.org/10.1093/poq/nfv014>
- Revilla, M., Couper, M. P., & Ochoa, C. (2019). Willingness of online panelists to perform additional tasks. *Methods, Data, Analyses*, 13(2), 223–252. <https://doi.org/10.12758/mda.2018.01>
- Sakshaug, J. (2020). Linking surveys with big data: Issues of consent. In B. Klumpe, J. Schröder, & M. Zwick (Eds.), *Qualität bei zusammengeführten Daten – Befragungsdaten, Administrative Daten, neue digitale Daten: Miteinander besser?* (pp. 163–173). Springer VS.
- Sakshaug, J., Couper, M. P., Ofstedal, M. B., & Weir, D. (2012). Linking survey and administrative records: Mechanisms of consent. *Sociological Methods & Research*, 41(4), 535–569. <https://doi.org/10.1177/0049124112460381>
- Sakshaug, J., Tutz, V., & Kreuter, F. (2013). Placement, wording, and interviewers: Identifying correlates of consent to link survey and administrative data. *Survey Research Methods*, 7(2), 133–144. <https://doi.org/10.18148/srm/2013.v7i2.5395>
- Sakshaug, J. W., & Kreuter, F. (2014). The effect of benefit wording on consent to link survey and administrative records in a web survey. *Public Opinion Quarterly*, 78(1), 166–176. <https://doi.org/10.1093/poq/nfu001>
- Sakshaug, J. W., Schmucker, A., Kreuter, F., Couper, M. P., & Singer, E. (2019). The effect of framing and placement on linkage consent. *Public Opinion Quarterly*, 83(S1), 289–308. <https://doi.org/10.1093/poq/nfz018>
- Sakshaug, J. W., Wolter, S., & Kreuter, F. (2015). Obtaining record linkage consent: Results from a wording experiment in Germany. *Survey Methods: Insights from the Field*, 12. <https://doi.org/10.13094/SMIF-2015-00012>
- Sala, E., Burton, J., & Knies, G. (2012). Correlates of obtaining informed consent to data linkage: Respondent, interview, and interviewer characteristics. *Sociological Methods & Research*, 41(3), 414–439. <https://doi.org/10.1177/0049124112457330>
- Sala, E., Knies, G., & Burton, J. (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(5), 455–473. <https://doi.org/10.1080/13645579.2014.899101>
- Silber, H., Breuer, J., Beuthner, C., Gummer, T., Keusch, F., Siegers, P., Stier, S., & Weiß, B. (2022). Linking surveys and digital trace data: Insights from two studies on determinants of data sharing behaviour. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(Supplement\_2), S387–S407. <https://doi.org/10.1111/rssa.12954>
- Singer, E. (2011). Toward a benefit-cost theory of survey participation: Evidence, further tests, and implications. *Journal of Official Statistics*, 27(2), 379–392. <https://www.scb.se/contentassets/f6bcee6f397c4fd68db6452fc9643e68/toward-a-benefit-cost-theory-of-survey-participation-evidence-further-tests-and-implications.pdf>
- Singer, E., & Ye, C. (2013). The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1), 112–141. <https://doi.org/10.1177/0002716212458082>
- Sloan, L., Jessop, C., Al Baghal, T., & Williams, M. (2020). Linking survey and Twitter data: Informed consent, disclosure, security, and archiving. *Journal of Empirical Research on Human Research Ethics*, 15(1–2), 63–76. <https://doi.org/10.1177/1556264619853447>
- Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, 38(5), 503–516. <https://doi.org/10.1177/0894439319843669>
- Stone, C., Noel, H., & Weir, D. (2013). *Obtaining administrative record linkage consent by mail: Examining the impact of incentives and telephone follow-up* [Paper presentation]. 68th Annual Conference of the American Association for Public Opinion Research, Boston, MA, United States.
- Walzenbach, S., Burton, J., Couper, M. P., Crossley, T. F., & Jäckle, A. (2022). Experiments on multiple requests for consent to data linkage in surveys. *Journal of Survey Statistics and Methodology*, 11(3), 518–540. Advance online publication. <https://doi.org/10.1093/jssam/smab053>
- Witte, N., Schaurer, I., Schröder, J., Décieux, J. P., & Ette, A. (2021). Enhancing participation in probability-based online panels: Two incentive experiments and their effects on response and panel recruitment. *Social Science Computer Review*. Advance online publication. <https://doi.org/10.1177/08944393211054939>