Open Access Repository
www.ssoar.info

Understanding the effects of conceptual and analytical choices on 'finding' the privacy paradox: A specification curve analysis of large-scale survey data

Masur, Philipp K.

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Masur, P. K. (2023). Understanding the effects of conceptual and analytical choices on 'finding' the privacy paradox: A specification curve analysis of large-scale survey data. *Information, Communication & Society*, *26*(3), 584-602. https://doi.org/10.1080/1369118X.2021.1963460

Nutzungsbedingungen:
Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de

Terms of use:
This document is made available under a CC BY-NC-ND Licence (Attribution-Non Comercial-NoDerivatives). For more Information see:
https://creativecommons.org/licenses/by-nc-nd/4.0

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz-Gemeinschaft

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-93005-8

Routledge
Taylor & Francis Group

Information, Communication & Society

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rics20

Understanding the effects of conceptual and analytical choices on ‘finding’ the privacy paradox: A specification curve analysis of large-scale survey data

Philipp K. Masur

To cite this article: Philipp K. Masur (2023) Understanding the effects of conceptual and analytical choices on ‘finding’ the privacy paradox: A specification curve analysis of large-scale survey data, Information, Communication & Society, 26:3, 584-602, DOI: 10.1080/1369118X.2021.1963460

To link to this article: https://doi.org/10.1080/1369118X.2021.1963460

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 06 Aug 2021.

Article views: 3202

View Crossmark data

View supplementary material

Submit your article to this journal

View related articles

Citing articles: 3 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rics20

INFORMATION, COMMUNICATION & SOCIETY
2023, VOL. 26, NO. 3, 584–602
https://doi.org/10.1080/1369118X.2021.1963460

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Understanding the effects of conceptual and analytical choices on ‘finding’ the privacy paradox: A specification curve analysis of large-scale survey data

Philipp K. Masur

Department of Communication Science, Vrije Universiteit Amsterdam (VU), Amsterdam, Netherlands

ABSTRACT
The privacy paradox suggests that privacy concerns do not relate to privacy-related behavior. Although it has inspired numerous studies, findings remain inconclusive. Some of the inconsistencies in published findings may be explained by a strong heterogeneity in the conceptual and analytical choices that researchers implement when investigating the privacy paradox. Based on representative survey data of the 27 EU member states (2011: $n = 8{,}962$; 2015: $n = 10{,}526$; 2019: $n = 11{,}428$), I investigated the effect of conceptual and analytical decisions on ‘finding’ the privacy paradox. Specification curve analyses revealed that the magnitude and statistical significance of the relationship between privacy concerns and information disclosure is contingent on the operationalization of the independent variable, the inclusion of covariates, and the age of the studied population. The relationship between online privacy concerns and using social media privacy settings, in contrast, was less influenced by analytical decisions. Yet, the relationship was stronger in younger people and increased over time. The findings call for more transparency in analyzing research data. Evaluating the implications of analytical choices will help to establish best practices and advance cumulative knowledge creation in privacy research.

ARTICLE HISTORY
Received 8 September 2020
Accepted 9 July 2021

KEYWORDS
Privacy paradox; privacy concerns; information disclosure; privacy protection; specification curve analysis; Eurobarometer

Much research on online privacy has been inspired by a continuously resurfacing phenomenon: the privacy paradox. When scholars first investigated the relationship between privacy concerns and privacy behaviors on social network sites (SNSs), they were puzzled to find that individuals with strong privacy concerns nonetheless disclosed a lot of private information (Acquisti & Grossklags, 2005; Spiekermann et al., 2001). Although the term *privacy paradox* was originally coined to describe differences between adults’ concerns and teenagers’ ignorance of the public nature of the Internet (Barnes, 2006), it has become synonymous with this concern-behavior discrepancy. If the privacy

CONTACT Philipp K. Masur p.k.masur@vu.nl Department of Communication Science, Vrije Universiteit Amsterdam (VU), De Boelelaan 1081/1081 HV, Amsterdam, Netherlands

Supplemental data for this article can be accessed online at https://doi.org/10.1080/1369118X.2021.1963460.

© 2021 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-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

paradox really exists, it suggests that people behave irrationally and fail to translate their own concerns into actions to protect their privacy. Understanding whether people's online behavior is indeed paradoxical is thus important as policy making, e-commerce, and platform design rest on assumptions about people's privacy concerns and privacy behaviors.

Many scholars have investigated the privacy paradox (more than 8,030 hits on Google Scholar[1]; for literature reviews, see Barth & de Jong, 2017; Kokolakis, 2017). Nonetheless, the findings remain inconclusive. Whereas some studies found support for the privacy paradox (e.g., Norberg et al., 2007; Taddei & Contena, 2013; Tufekci, 2008), others did not (e.g., Awad & Krishnan, 2006; Dienlin & Trepte, 2015; Krasnova et al., 2010). A first meta-analysis (Baruh et al., 2017) suggests a negative, but small relationship between privacy concerns and information disclosure ($r = -.13$, 95% [−.18, −.07]; 37 studies) and a positive, slightly stronger relationship between privacy concerns and privacy protective behaviors ($r = .17$, 95% [.12, .23]; 35 studies).

Despite this meta-analytical finding, the privacy paradox continues to be debated among privacy researchers (e.g., Barth et al., 2019) and the public alike (e.g., Naughton, 2019). With the increasing number of studies, the sheer volume of collected data, and the various analytical approaches used to study the privacy paradox, it seems that our understanding of the actual phenomenon has become blurred. Problematic publication practices such as reinventing the wheel, unawareness of similar work, and conceptual instead of methodological replications contribute to the accumulation of incompatible findings. Despite promising advances in the field, it is time to take stock and assess potential weaknesses in seemingly cumulative evidence for or against the privacy paradox.

This paper aims to understand inconsistencies in research on the privacy paradox by investigating the impact of conceptual and analytical choices on 'finding' the privacy paradox. Therefore, I first discuss researchers' degrees of freedom as a potential cause of heterogeneity in results. I then review the literature on the privacy paradox to identify prominent theoretical, conceptual, and analytical choices. Finally, I report and discuss results from two specification curve analyses that remodel these choices and quantify their effects on the relationship between privacy concerns and behavior based on three large-scale survey data sets.

Multiverse or specification curve analysis approaches

Variance in obtained results from studies that test a similar hypothesis may partly stem from researchers' degrees of freedom when collecting and analyzing their data (Gelman & Loken, 2013). *Specification curve* or *multiverse analyses* (Simonsohn et al., 2020; Steegen et al., 2016) start from the observation that 'data used in an analysis are […] not just passively recorded in an experiment or an observational study. Rather, data are to a certain extent actively constructed' (Steegen et al., 2016, p. 702). Researchers decide on specific operationalizations in line with their theoretical assumptions, define the population from which they will sample, and transform raw data to be ready for the analysis. Processing steps often include selecting items, transforming variables, including control variables, creating subsets, and choosing an estimation approach. In each step, a researcher usually has several options to choose from. In an ideal setting, one chooses the best option based on sound and justifiable arguments.

In many research areas – including research on the privacy paradox – such decisions vary considerably and are often arbitrary or equally defensible. Sometimes, clear justifications are lacking completely. Yet, each decision leads to a different dataset and a multiverse of statistical results. Multiverse or specification curve analyses aim to estimate all reasonable specifications that are consistent with the underlying theory, are expected to be statistically valid, and are not redundant with other specifications (Simonsohn et al., 2020). Their potential lies in exposing the impact of hidden degrees of freedom and, in some cases, in obtaining more robust estimates of an effect of interest.

However, if the specifications are not truly arbitrary, i.e., an alternative is objectively justified over another, such analyses can produce misleading results and hide 'meaningful effects within a mass of poorly justified alternatives' (Del Giudice & Gangestad, 2021, p. 1). Thus, it is important to distinguish between different types of decisions (for a formal differentiation, see Del Giudice & Gangestad, 2021, pp. 11–12): First, *type E decisions* (principled equivalence) refer to instances in which empirical evidence or theoretical considerations indicate that alternative analyses are equivalent. For example, such decisions may include deciding between similar measures with comparable validity and reliability or deciding between arbitrary thresholds to exclude outliers. Second, *type N decisions* (principled non-equivalence) refer to instances in which available evidence or theoretical considerations suggest that alternatives are non-equivalent, for example, when deciding between conceptually related, yet different measures, deciding between theoretically motivated subsets (e.g., males vs. females, adolescents vs. adults), or deciding to include potentially relevant covariates. Although such decisions should be justified by theory, nonetheless, studies vary considerably with regard to how variables are operationalized, measured, and estimated, which control variables are included, and what population is investigated. Instead of relying on theory, such decisions are often based on practicability, accessibility, or even financial restrictions (e.g., length of questionnaire or access to certain population). Nevertheless, results from such unequal studies are often interpreted as support for or against one broad hypothesis (e.g., the privacy paradox). Finally, there is often insufficient information about whether a decision is truly arbitrary or not. Such *type U decisions* (uncertain decisions) include, for example, when two measures seem conceptually similar, but there is no empirical evidence for comparable validity, or when the influence of potential confounders on the effect size of interest has not yet been investigated.

According to several scholars (Del Giudice & Gangestad, 2021; Simonsohn et al., 2020), including decisions that are not truly arbitrary (i.e., type N and type U decisions) in a multiverse analysis can produce bias, exaggerate the perceived exhaustiveness, and thereby reduce its informative power. That said, multiverse analyses that do include type N and type U decisions may help to understand how these different types of decisions affect an outcome of interest. In such cases, however, multiverse analyses should be deliberately exploratory and alternatives should be examined separately (Del Giudice & Gangestad, 2021; Simonsohn et al., 2020). Recent implementations of a such multiverse analyses have shown that obtaining a statistically significant result (which is often the primary basis for rejecting or supporting a hypothesis) and, more importantly, the effect's magnitude can heavily depend on the three types of decisions or combinations thereof (e.g., Orben & Przybylski, 2019; Rohrer et al., 2017; Steegen et al., 2016).

Conceptual and analytical decisions in the privacy paradox literature

To inform the present specification curve analyses, I reviewed the extensive literature on the privacy paradox. Table A6 in the online supplement (https://osf.io/v85xf/) provides an overview of 30 exemplary studies (including the most influential studies, based on citations) that explicitly or implicitly tested the privacy paradox. Although the investigated hypotheses are comparable across almost all studies (variants of 'privacy concerns are positively related to privacy behaviors' or 'privacy concerns are negatively related to self-disclosure'; note that the paradox refers to the observation that these hypotheses are *not* supported), studies differ considerably with regard to all three types of decisions.

In a first step, it is important to note that studies investigating the privacy paradox adopted different theoretical frameworks. Most scholars relied on socio-psychological theories of privacy (cf. Bazarova & Masur, 2020). Primarily viewing privacy as withdrawal from social interactions (Westin, 1967), such approaches often argue that people engage in a constant optimization processes to balance their desired and achieved level of privacy (Altman, 1976). Here, privacy concerns are regarded as a response to an undesired level of privacy, and privacy behaviors (such as minimizing information disclosure or engaging in privacy protection) are means to achieve the desired level of privacy. A similar theoretical rationale is employed by the privacy calculus literature (e.g., Dienlin & Metzger, 2016; Krasnova et al., 2010), which posits that individuals weigh the risks (expressed in privacy concerns) and benefits (e.g., assumed gratifications) before disclosing personal information.

Recent frameworks emphasize that individual control over personal information is challenging in online environments (Trepte, 2020). Building on Altman's work, newer approaches acknowledge that online privacy management requires collective actions based on rule-based negotiations of boundaries and ownership (De Wolf, 2020; Marwick & boyd, 2014; Petronio, 2002). Although less prominent in the literature on the privacy paradox, such networked approaches could explain why individual concerns may not necessarily lead to individual behavioral responses. Other approaches propose that claims to privacy are claims to an appropriate flow of information that is contingent on situational (Masur, 2018) and contextual circumstances (Nissenbaum, 2010). Differentiations between horizontal (i.e., with regard to other users) and vertical privacy levels (i.e., with regard to online service providers and institutions) further help to identify boundary conditions of feasible privacy management (Epstein & Quinn, 2020).

Distinguishing these levels of theoretical analysis has implications for various conceptual decisions and also for detecting the effect of interest. Although most research on the privacy paradox adopts an economic model of privacy, in which personal information is regarded as an exchange value, the following discussion shows that different interpretations can have strong implications for conceptual and analytical decisions.

A first conceptual decision – a type N decision and sometimes a type U decision – refers to the operationalization of the independent (privacy concerns) and dependent variable (privacy behavior). For example, whereas early investigations often focused on predicting what type of information users disclosed in public profiles (e.g., Acquisti & Gross, 2006; Tufekci, 2008), later studies often investigated more general measures of self-disclosure (e.g., Dienlin & Metzger, 2016; Hallam & Zanella, 2017; Krasnova et al., 2010; Taddicken, 2014). Other studies focused on privacy management strategies, such as whether or not users restricted access to their SNS profile (Chen, 2018; Chen &

Chen, 2015; Utz & Krämer, 2009). Whereas most studies used self-reports, some also used log data (Acquisti & Gross, 2006; Reynolds et al., 2011) or measured behavioral intentions (Hallam & Zanella, 2017; Trepte et al., 2017). Table A6 shows that different operationalizations of the dependent variable may yield different or even contrary results. Among other notable differences, privacy concerns seem to be stronger predictors of privacy protection behaviors than of the amount of self-disclosure.

Early studies often tried to predict specific behaviors (e.g., using the real name on Facebook) with rather general privacy concerns (e.g., 'How concerned are you about your online privacy'), often resulting in the unexpected, paradoxical relationship (e.g., Acquisti & Gross, 2006; Tufekci, 2008). Designing items by following the *principle of compatibility* (i.e., items that comply in terms of action, target, context, and time; Fishbein & Ajzen, 2010), in contrast, often yielded significant relationships between privacy concerns and various privacy behaviors (Dienlin & Trepte, 2015). Differences also emerged if scholars differentiated vertical privacy concerns and horizontal privacy concerns (Masur, 2018; Walrave et al., 2012), or when they measured attitudes or perceived risks instead of privacy concerns (Dienlin & Trepte, 2015; Krasnova et al., 2010).

A second decision refers to the inclusion of control variables. Whereas some studies did not include control variables (e.g., Acquisti & Gross, 2006), others controlled for attitudes and intentions (e.g., Dienlin & Trepte, 2015), trust towards providers (e.g., Utz & Krämer, 2009), benefits of using online services (e.g., Dienlin & Metzger, 2016; Krasnova et al., 2012), privacy cynicism (Lutz et al., 2020), self-efficacy (Chen & Chen, 2015; Dienlin & Metzger, 2016), or privacy violation experiences (e.g., Awad & Krishnan, 2006). Many studies also controlled for socio-demographics (e.g., age, gender, education; Blank et al., 2014; Dienlin et al., 2019; Lutz & Strathoff, 2014). However, it is difficult to discern whether the inclusion of controls systematically affects the outcome. Theories are often vague as to whether such variables are confounders (which should be controlled for) or potential mediators (which should not be controlled for) of the effect of privacy concerns on privacy behaviors (cf. Dienlin & Trepte, 2015). Hence, the inclusion of control variables must be regarded as a type U decision.

Third, studies often focused on different contexts, platforms, and populations (see Table A6). For example, many studies were based on student samples (e.g., Acquisti & Grossklags, 2005; Barnes, 2006; Norberg et al., 2007; Tufekci, 2008), some on representative samples, but culturally diverse populations (e.g., Blank et al., 2014; Lutz & Strathoff, 2014; Taddicken, 2014), on users of specific applications (Egelman et al., 2013), or on Internet users in general (e.g., Dienlin et al., 2019; Lutz & Strathoff, 2014). Whereas most studies explicitly focused on Facebook users, some explore alternative platforms such as Hyves (Utz & Krämer, 2009) or culture-specific SNSs (Trepte et al., 2017). Although robust evidence that could guide justified decisions in this regard is missing, deciding between populations must be regarded as a type N decision as several scholars have pointed out that privacy-related phenomena are affected by structural markers such as age or education (Madden & Rainie, 2015).

Finally, Table A6 also shows that existing studies differ considerably regarding how variables are estimated (e.g., mean scores vs. latent modeling) and model estimation approaches (e.g., analysis of variance, multiple regression, structural equation models …). Other types of analytical decisions (e.g., filtering decisions, imputation, item selection procedures, etc.) are often not reported, but could add another layer of heterogeneity.

Systematic variability due to such type E decisions, however, is less obvious because fully transparent reports of the complete analytical procedure are often not available.

The present study

Based on this review of the literature, this study systematically investigated the effect of conceptual and analytical choices on 'finding' the privacy paradox based on three large-scale, representative surveys of the European Union (overall $N = 82{,}078$).[2] These datasets include various items that could be used to analyze the privacy paradox and to recreate some of the identified conceptual and analytical decisions (see Table 1). The first analysis is based on data collected in 2011 and investigated the relationship between *privacy concerns* and *information disclosure* on SNS. The data collected in 2015 and 2019, in contrast, were used to investigate the relationship between *privacy concerns* and the *use of privacy settings* at two different points in time (both surveys implemented the same items).

Analysis 1: online privacy concerns and information disclosure

Sample

I used the data of the Special Eurobarometer 359 which was part of the general Eurobarometer 74.3. The data collection took place between November 25 and

Table 1. Overview of the conceptual and analytical choices (specifications).

Decisions	Analysis 1 (Eurobarometer 2011)	Analysis 2 (Eurobarometer 2015 and 2019)
Operationalization of the dependent variable (y)	Sum score of all 14 binary items measuring information disclosure[a]	Binary variable measuring privacy setting use (1 = changed default settings; 0 = has not changed)[e]
Operationalization of the independent variable (x)	Three items measuring different types of privacy concerns (4-point scale), the mean of all three items[b]	Single item measuring privacy concerns (4-point scale)[f]
Model estimation	Linear multilevel regression models (dependent variable was roughly normally distributed)	Logistic multilevel regression models (binary outcome)
Inclusion of covariates (controls)	No covariates, each covariate individually, all covariates[c]	No covariates, each covariate individually, all covariates[g]
Age-based subsets (age)	≤32 years, >32 years (median split), all participants[d]	≤41 years, >41 years (median split), all participants[d]
Time of data collection (year)	2011	Either 2015, 2019, and both
Number of specifications	108	90

[a] Similarly operationalized by e.g., Taddicken (2014), Dienlin et al. (2019), Tufekci (2008) and others (see Table A6, 'dependent variable').

[b] All three items measure a type of vertical privacy concern and are thus comparable to the measures of e.g., Chen (2018), Dienlin et al. (2019), Hallam and Zanella (2017), and others (see Table A6, 'independent variable').

[c] Control variables include perceived control, attitude toward information sharing, use of privacy settings, Internet use at home and at work, gender, and age. Similar control variables have been used in the literature (see Table A6, 'inclusion of control variables').

[d] To compare different age-groups as prior results seem to differ depending on whether it focused on student samples or the larger population (see Table A6, 'demographic structure').

[e] Similarly operationalized by e.g., Blank et al. (2014), Chen and Chen (2015), Masur (2018) and others (see Table A6, 'dependent variable').

[f] Concern about not having complete control over what one has disclosed online, comparable to measures by Chen (2018) or Dinev & Hart (2006).

[g] Control variable include perceived control, subjective knowledge about privacy settings, internet use at home and at work. Similar control variables have been used throughout the literature (see Table A6, 'inclusion of control variables').

December 17, 2010. The data are representative for the population of all 27 EU member states (aged 15 years and older). Information about the sampling procedure can be found in the annexes of the Special Eurobarometer 359 report (European Commission, 2011).

Overall, $N = 26{,}574$ participants were interviewed (face-to-face or computer-assisted personal interviews in the appropriate languages). For the following specification curve analysis, I only included participants who indicated they use SNSs (33.7%) and who provided answers to the information disclosure items ($n = 8{,}962$). These participants were $M = 34.6$ years on average ($SD = 13.93$, *range* = 15–93) and 52.8% were female. Samples sizes varied across analyses due to sub group analyses (e.g., younger vs. older participants; see Figure 1C) and – to a smaller degree – due to missing values in all variables (median $n = 4{,}402$; *min* = 3,813, *max* = 8,865).

Measures

Information disclosure

Participants were asked which out of 14 items (e.g., medical information, name, home address …) they have already disclosed when registering or using SNSs. Answer options were 0 = *No* and 1 = *Yes*. The overall amount of information disclosure can be understood as a formative concept. By presenting participants an exhaustive list of items and asking them which they have disclosed, we can estimate the overall level of information disclosure by creating the sum score of all 14 items. Higher values thus represent a higher amount of disclosed information items ($M = 4.58$, $SD = 2.50$).[3]

Online privacy concerns

The dataset included three items that measured online privacy concerns. The first referred to concerns about online behavior being recorded and was measured on a scale ranging from 1 = *not at all concerned* to 4 = *very concerned* ($M = 2.42$, $SD = 0.87$). The second item represented concerns about unwanted use of personal information. The scale ranged from 1 = *very uncomfortable* to 4 = *very comfortable* ($M = 2.82$, $SD = 0.81$). The last item referred to concerns about targeted advertising and personalization and was measured on a scale ranging from 1 = *not at all concerned* to 4 = *very concerned* ($M = 2.62$, $SD = 0.81$).

Control variables

I included eight control variables that prior research has identified as being related to information disclosure (see Table 1 and A6): whether or not people changed the default privacy settings (0 = *no*, 1 = *yes*; 50% changed their privacy settings), their perceived control over the use of their personal information (1 = *no control* to 3 = *complete control*), Internet use (1 = *Never* to 6 = *Everyday/Almost everyday*) at home and at work, people's attitudes towards sharing information ('Disclosing personal information is not a big issue for you'; 1 = *totally disagree* to 4 = *totally agree*), gender and age.

Data analysis

In the first step, I investigated which of the typically implemented decisions in the published literature (Table A6) could be recreated within the limits of the data set (Table 1). Analytical choices included the operationalization of the independent variable *online privacy concerns* (three single items, their mean score; 4 choices), the inclusion of control variables (no covariates, each covariate individually, all covariates; 9 choices), and the creation of age-based subsets (all, younger than 32 years (median); older than 32 years; 3 choices), resulting in 108 specifications.

Second, I estimated linear multilevel regression models for each specification as participants were nested in countries (4.4% of information disclosure was explained by between-country differences; ICC = .044). I used listwise deletion for missing data. I then extracted the standardized regression coefficient and confidence intervals from each output. Figure 1 provides an overview of how the coefficients are distributed (the specification curve, upper panel) and how analytical choices affect the magnitude and statistical significance of this effect (lower panel). Simonsohn et al. (2020) propose a bootstrapping approach to test whether, when considering all possible specifications, the results are inconsistent with the results when the null hypothesis is true. A requirement for this step is that all analytical decisions are indeed arbitrary. As emphasized earlier, the following analyses included type E and type U decisions. Hence, I decided against this third step.

Instead, I investigated which of the analytical decisions explained most variance in the effect sizes. I therefore included all choices as random effects in a multilevel model without predictors. Based on the random effect variance, I computed intra-class correlation coefficients (ICC) that quantify the amount of variance in the specification curve that is attributable to respective decisions (including their interactions). The distribution of variance components is visualized in Figure 3.

Results

Across all specifications, 78.7% resulted in a significant negative relationship between privacy concerns and information disclosure (Figure 1A). The median association of privacy concerns and information disclosure was $\beta = -.051$ (median absolute deviation (*Mad*) = 0.026, min = –.098, max = –.005, median $b = -0.157$; median $n = 4{,}402$). Overall, the effects were small at best (Cohen, 1988; r = .10 represents a *small* effect). However, a quarter of the effect sizes (23.1%) were within the confidence intervals of the meta-analytical finding of Baruh et al. (2017; r = –.13, 95% CIs [–.18, –.07]). Most of the variance in the obtained effects was explained by different operationalizations of the independent variable (54.9%; see Figure 3, *left*), the inclusions of different control variables (7.8%), different age groups (3.9%) as well as the interaction between age group and independent variable operationalization (26.3%).

The analysis revealed that using *concerns about unwanted use of information* as the independent variable yielded less diverse and mostly significant negative effect sizes (median $\beta = -0.051$, *Mad* = .007, min = –.067, max = –.029, median $b = -0.155$). *Concerns about targeted advertisement and personalization* produced slightly larger, but also more diverse estimates (median $\beta = -.068$, *Mad* = .381, min = –.098, max = –.022, median $b = -0.203$). Here the interaction with age becomes apparent: For older participants, concerns

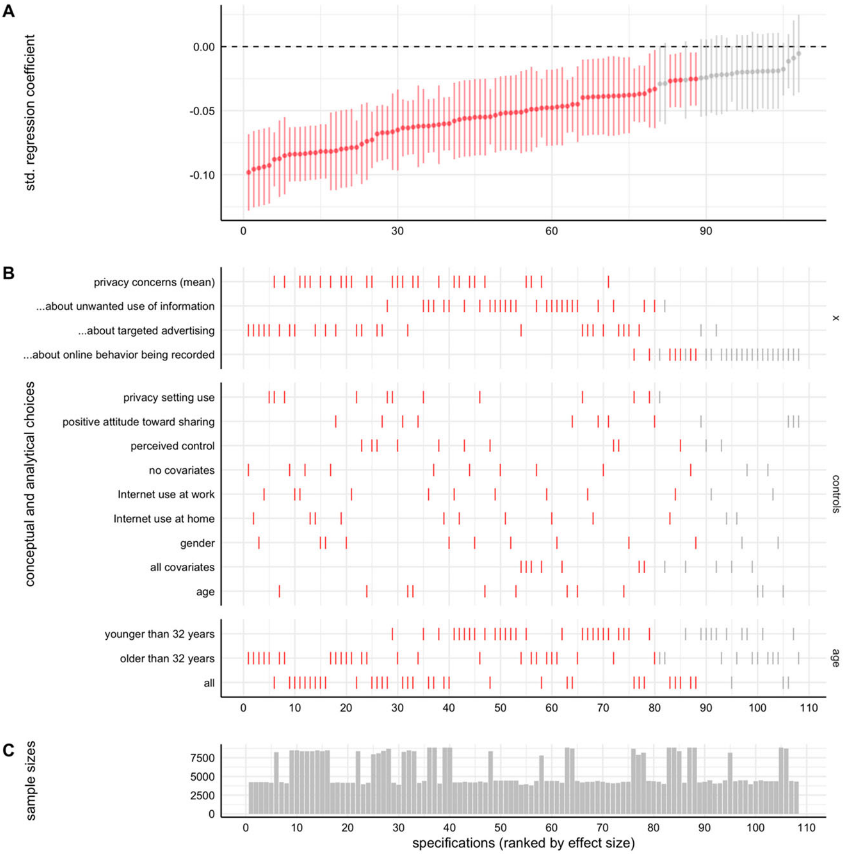

Figure 1. Specification curve analysis of the relationship between privacy concerns and information disclosure (108 specifications). The unstandardized coefficients (and their 95% CIs) are shown in the upper panel (A). The middle panel (B) depicts how specific conceptual and analytical choices affected the resulting coefficient. The lower panel (C) shows the sample sizes for each specification (red: $p < .05$; gray: $p > .05$).

about targeting and personalization were more strongly related to information disclosure (median $\beta = -.093$, *Mad* = .001, min = –.098, max = –.051, median $b = -0.273$) compared to younger participants (median $\beta = -.038$, *Mad* = .001, min = –.040, max = –.022, median $b = -0.118$). Considerably smaller and mostly non-significant effect sizes were obtained if *concerns about online behavior being recorded* was used as predictor (median $\beta = -.021$, *Mad* = .004, min = –.037, max = –.005, median $b = -0.058$*)*.

The impact of control variables is somewhat difficult to assess as results related to any of the covariates span the whole specification curve. In general, including all control variables produced some of the smallest effect sizes (median $\beta = -.037$, *Mad* = .017, min = –.051, max = –.012, median b = –0.107; Figure 1B). However, including privacy setting use increased the effect sizes considerably (median $\beta = -.064$, *Mad* = .033, min = –.093, max = –.029, median $b = -0.195$). Furthermore, focusing on younger participants produced smaller effect sizes (median $\beta = -.039$, *Mad* = .022, min = –.066, max = –.009, median b = –0.122) compared to older participants (median $\beta = -.050$, *Mad* = .043, min = –.098, max = –.005, median b = –0.158).

Overall, the smallest effect size was obtained if *positive attitude toward sharing information* was controlled for, *concerns about behavior being recorded* was used as the independent variable, and only younger participants were analyzed ($b = -0.015$, $\beta = -.005$, n.s.). The largest effect size was obtained when all participants were analyzed, the composite mean score was used as independent variable, and privacy setting use was controlled for ($b = -0.289$, $\beta = -.098$, $p < .001$).

Analysis 2: online privacy concerns and privacy setting use

Sample

For this analysis, I used the data of the Special Eurobarometer 431, which was part of the general Eurobarometer 83.1, and the Special Eurobarometer 487a, which was part of the general Eurobarometer 91.2. The data collection for the former took place between February 28 and March 9, 2015 and for the latter between March 15 and 29, 2019. Both datasets are representative for the population of all 27 EU member states (aged 15 years and older). Exact information about the sampling procedure can be found in the annexes of the Special Eurobarometer 431 (European Commission, 2015) and the Special Eurobarometer 487a (European Commission, 2019b).

Overall, $N = 27{,}980$ participants were interviewed in 2015 and $N = 27{,}524$ participants in 2019 (face-to-face or computer-assisted personal interviews in appropriate languages). For the specification curve analysis, I only included participants who indicated they use SNS (2015: 78.3%; 2019: 85.6%) and provided answers to the item measuring privacy setting use on SNSs (overall $n = 21{,}954$; $n_{2015} = 10{,}526$; $n_{2019} = 11{,}428$). In 2015, the average age was $M = 40.6$ years ($SD = 15.3$, $range = 15–98$) and 53.7% of the sample was female. In 2019, the average age was $M = 43.4$ years ($SD = 15.7$, $range = 15–92$) and 55.0% of the sample was female. Again, samples sizes varied across the analyses due to subgroup analyses (see Figure 2C) and missing values in all variables (median $n = 10{,}350$; $min = 4{,}430$, $max = 21{,}954$).

Measures

Use of privacy settings

Whether participants changed privacy settings was assessed with a single item ('Have you ever tried to change the privacy settings of your personal profile from the default settings on an online social network?'). Answer options were 0 = *no* and 1 = *yes.* In 2015, 55.0% of the participants indicated they changed their privacy settings. In 2019, slightly fewer participants changed the default SNS privacy settings (54.2%).

Online privacy concerns. In contrast to the Eurobarometer survey from 2011, the surveys from 2015 and 2019 included only a single item to measure privacy concerns ('How concerned are you about not having complete control over the information you provide online?'). Answer options ranged from 1 = *not at all concerned* to 4 = *very concerned* ($M_{2015} = 2.75$, $SD = 0.76$; $M_{2019} = 2.67$, $SD = 0.79$).

Control variables

Again, I included control variables that were found to be related to privacy protection behavior in prior studies (see Table 1). These included perceived control over personal information (1 = *no control at all* to 3 = *complete control*), subjectively perceived difficulty of changing privacy settings ('How easy or difficult was it in your opinion to change the data protection settings on your personal profile?'; 1 = *very easy* to 4 = *very difficult*), reading an online service provider's privacy statement ('When you think about privacy policies on the Internet, which of the following sentences describes what you usually do?'; 1 = *You don't read them at all* to 3 = *You read them completely*), frequency of Internet use at home, Internet use at work, and SNS use (all three were measured on a scale ranging from 1 = *Never* to 6 = *Everyday/Almost everyday*), as well as gender and age.

Data analysis

Analytical choices were again identified based on the existing literature. For the following analyses, however, available options were somewhat limited as only few items could be used to test the privacy paradox in 2015 and 2019 (Table 1). Nonetheless, conceptual and analytical choices encompassed the inclusion of control variables (no covariates, each covariate individually, all covariates; 10 choices), and the creation of subsets (9 choices; age-based subsets and year), resulting in 90 specifications.

Due to the binary dependent variable, logistic multilevel models were estimated for each specification as participants were nested in countries (ICC = .045). I used listwise deletion for missing data and extracted the relevant coefficient (Odds Ratios; the exponentiation of the unstandardized coefficient obtained in the logistic regression) and confidence intervals from each output. The results are visualized in Figure 2. In a third step, I again investigated the variance composition of the specification curve (Figure 3).

Results

All specifications resulted in a positive and significant relationship between privacy concerns and the use of SNS privacy settings (median Odds Ratio = 1.258, *Mad* = 0.074; min = 1.134, max = 1.414; median *n* = 10,350). An OR of 1.258 indicates that a one-unit change in the independent variables (here privacy concerns on a 4-point scale) multiplies the chance of the outcome (i.e., a participant having changed the privacy settings) by a factor of 1.258.

Overall, the resulting effect sizes were quite homogenous and ranged from tiny to small at best (Cohen, 1988; OR = 1.4 represents a small effect of *r* = .10). None of the effect sizes were within the confidence intervals of the meta-analytical findings of Baruh et al. (2017); r = .17, 95% Cis [.12, .23]. However, in comparison to the first analysis, effect sizes were slightly larger, with 45.6% being above *r* = .07 (lower CI of the relationship between privacy concerns and information disclosure based on Baruh et al., 2017).

Similar to the first analysis, including control variables did not account for much variance in the obtained effect sizes (9.1%, Figure 3, *right*). Most variance was explained by the year in which the data was collected (45.8%) and whether younger or older participants were analyzed (38.2%). A small amount of variance was also explained by the interaction between age and inclusion of controls (4.0%).

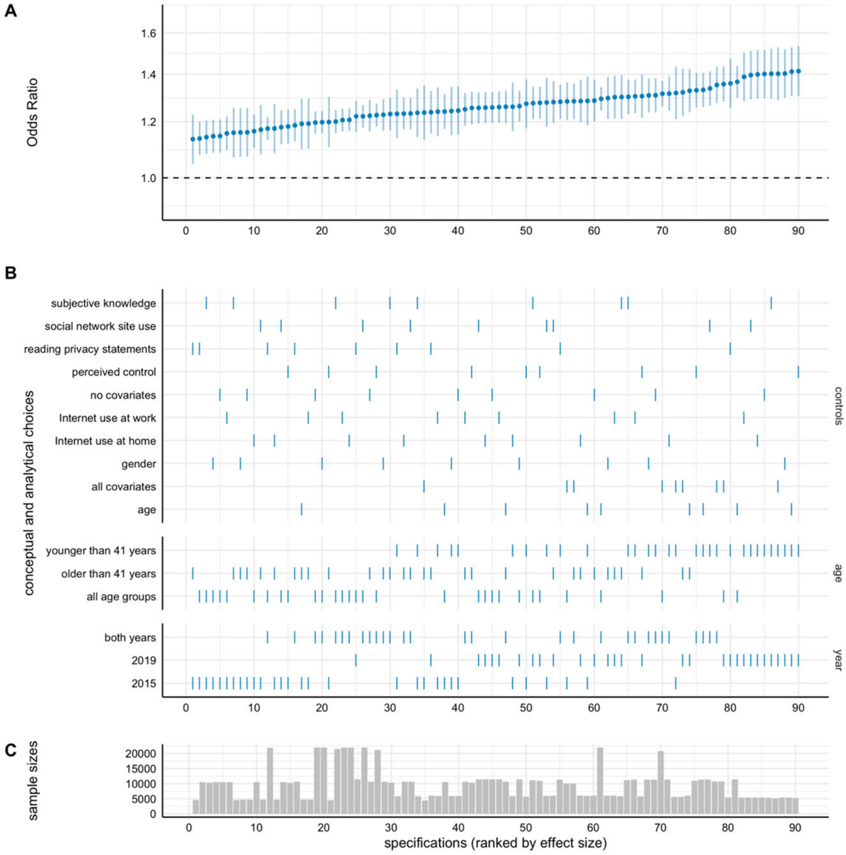

Figure 2. Specification curve analysis of the relationship between privacy concerns and privacy settings use (90 specifications). Odds ratios (and their 95% CIs) are shown in the upper panel (A). The middle panel (B) depicts the impact of conceptual and analytical choices. The lower panel (C) shows sample sizes for each specification (gray: $p > .05$; blue: $p < .05$).

The analysis showed that including different control variables did not affect the results considerably. Again, including all covariates produced some of the largest effect sizes (median OR = 1.317, *Mad* = 0.052, min = 1.236, max = 1.402), but specifications involving any of the covariates produced effect sizes across the whole specification curve range.

Most notably, the relationship between concerns and privacy setting use increased from 2015 (median OR = 1.192, *Mad* = 0.067, min = 1.134, max = 1.317) to 2019 (median OR = 1.302, *Mad* = 0.069, min = 1.221, max = 1.414; Figure 2B, *bottom*). Differentiating younger and older participants likewise produced considerable variance, with older participants yielding slightly smaller effect sizes (median OR = 1.234, *Mad* = 0.071, min = 1.134, max = 1.327; Figure 2B) than younger participants (median OR = 1.316, *Mad* = 0.074, min = 1.231, max = 1.414).

The smallest effect size was obtained for older participants in 2015 when reading privacy policies was controlled for (OR = 1.134, $p < .001$). The largest effect size was obtained by investigating only younger participants in 2019 and controlling for perceived control (OR = 1.414, $p < .001$).

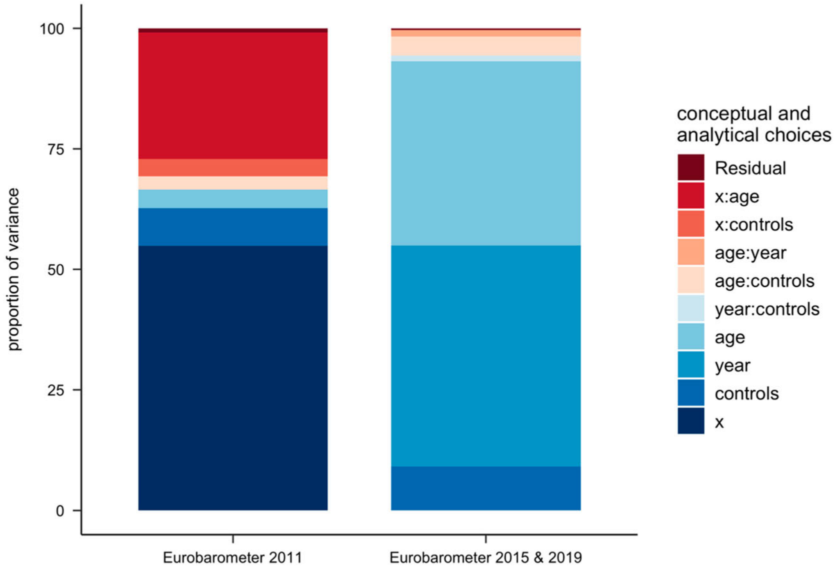

Figure 3. Distribution of variance components in the specification curve. The stacked bars represent how much variance in the specification curve is attributable to a specific conceptual or analytical choice (or their combinations). For example, the choice of a particular independent variable (x, the different measures of privacy concerns) is more influential on the outcome than including or excluding control variables (controls) in the analysis of the 2011 data set.

Discussion

The privacy paradox has sparked a fruitful line of research that helps to understand how privacy concerns relate to privacy behaviors. To contribute to recent attempts to map and summarize the field with systematic reviews (e.g., Barth & de Jong, 2017; Kokolakis, 2017) and meta-analyses (e.g., Baruh et al., 2017), this paper provides insights into the impact of researchers' degrees of freedom in designing and analyzing studies on obtaining a certain result.

The analyses reveal how comparatively simple conceptual and analytical choices lead to a multiverse of results that, in turn, may lead to vastly different conclusions. If we only focus on statistical significance, the analyses suggest that most specifications lead to the rejection of the privacy paradox. However, if we look at the actual effect sizes of these relationships, the picture is less clear. Only comparatively few specifications in both analyses yielded effect sizes that were *small* at best (cf. Cohen, 1988).

The analyses show that differences in item wording or the inclusion of control variables matters in whether we obtain a practically relevant effect size. For example, privacy concern that focused explicitly on personal information and thus followed the principle of comparability (e.g., concerns about personalization based on disclosed information) were better predictors for information disclosure than general concerns (e.g., concerns about online behavior being recorded). Such a decision is not arbitrary (type N decision) and may partly explain why studies who implemented different operationalizations of privacy concerns to test the same hypothesis yielded different results.

The findings further suggest that we should not expect to find similar results for different behaviors. For example, it may not be paradoxical that people with strong concerns still want to participate in online communication and thus share private information, while simultaneously trying to protect their privacy in other ways. The second analysis revealed that all specifications resulted in a significant and positive relationship between privacy concerns and privacy setting use. In other words, even a researcher's flexibility in analyzing this dataset could not lead to a finding consistent with the privacy paradox (if only evaluated based on significance). Privacy concerns seem to be slightly better predictors of privacy setting use (an active privacy protection strategy) than reducing information disclosure (a form of withdrawal that would lead to a considerable loss of disclosure-related benefits) – a finding that is consistent with the newer literature (Baruh et al., 2017).

Overall, the inclusion of control variables had a less consistent impact on results. However, the analyses showed that certain variables could be particularly important to control for. For example, controlling for privacy settings use yielded larger effect sizes for the relationship between privacy concerns and information disclosure. Similarly, perceived knowledge and whether or not someone usually reads privacy policies seem to be connected to privacy setting use. Furthermore, perceived control (similar to self-efficacy) seems generally worthwhile to investigate when analyzing the privacy paradox.

It is further notable that the link between concerns and privacy setting use became stronger over time. This is surprising in so far as neither overall privacy concerns nor privacy setting use increased from 2015 to 2019. One explanation could be a growing literacy among SNS users that helps to translate originally vague concerns into more reasonable privacy behavior. This begs the question of whether the relationship between privacy concerns and self-disclosure – which was only analyzed based on data from 2011 – likewise became stronger over time.

Finally, the analyses revealed an interesting age difference. On the one hand, the relationship between privacy concerns and information disclosure was stronger for older participants compared to younger participants. Based on classical effect size conventions, the present analysis even suggests that concerns predict privacy behaviors in older, but not in younger participants. On the other hand, the relationship between privacy concerns and privacy setting use was consistently stronger for younger participants compared to older participants. One explanation could be age-related differences in online privacy literacy (Masur, 2020) or the perceived importance of benefits of social media use (Dienlin & Metzger, 2016; Krasnova et al., 2010).

Limitations

First, although the analyzed datasets were comparatively large, the instruments used in the surveys were mostly single items. Paired with having few answer options, it is questionable how well such items capture variance in both concerns and behaviors. For this reason, this paper provides no final answer to the question of whether the privacy paradox exists. Other studies that have used more reliable instruments and implemented more sophisticated analysis methods (e.g., Dienlin et al., 2019) or tracked behavior instead (e.g., Nosko et al., 2010) should provide more robust evidence.

Second, the likelihood of obtaining a statistically significant effect also depends on the sample size. Different analytical choices or combinations thereof resulted in different sample sizes (Figure 1 and 2, *bottom*) and thus either reduce the power of finding a statistically significant effect or render even tiny effects significant. One should be careful in comparing statistical significance across specifications.

Finally, the present analysis is based on cross-sectional survey data. Although most research on the privacy paradox implemented such a research design, longitudinal panel designs, experiments, or qualitative interview studies may be less likely to yield as heterogeneous results.

Conclusion and future perspectives

The findings of this study suggest that privacy concerns are most likely related to both information disclosure and privacy setting use, but the relationship is small at best – at times even practically negligible. We need to acknowledge that concerns may generally matter, but to a smaller agree than often assumed. The present analysis has shown that whether or not the relationship is of practical relevance also depends to a certain degree on conceptual and analytical choices. To advance research on the privacy paradox, we should thus aim to create best practices based on more in-depth insights into the heterogeneity of privacy research. The following solutions are proposed:

First, future research should not expect to find consistent results that either refute or support the privacy paradox if factors such as the studied population (e.g., young vs. old), the type of privacy concerns (general vs. specific), the type of behaviors (withdrawal vs. protection), as well as inclusion of potential confounders or mediators vary. The present analyses suggest that good alignment between concerns and studied behaviors may be particularly important to estimate meaningful relations between such concepts. Items measuring privacy concerns should be carefully adapted to the context of each study as well as to the specific privacy behavior that one is interested in predicting.

Second, future researchers should not only focus on statistical significance, but rather ask what effect size is of practical relevance for a particular context or population. For example, the present findings revealed that the relationship between concerns about targeted advertisement and information disclosure among older people is small ($\beta = -.098$, no covariates), but the unstandardized coefficient ($b = -0.29$) suggests that an older person with low concerns (scoring 1 on the 4-point scale) discloses on average one more point of information ($\Delta b\ = 1.20$) than a person with strong concerns (4 on the 4-point scale). Such a difference could mean that an older, highly concerned person does not reveal their financial information online, whereas an older, hardly concerned person does – a relevant finding to understand why certain older people engage in online banking while others do not.

If the relationship between privacy concerns and privacy concerns is indeed negligible in a particular context or population, we should refrain from denoting it as ‘paradoxical’ and inquire about reasons for this discrepancy instead. Among others, person-related factors such as perceived control, self-efficacy and low privacy literacy (Masur, 2020), privacy cynicism (Lutz et al., 2020), or heuristic-decision making processes (Gambino et al., 2016) could explain surprisingly small effect sizes.

Finally, scholars should review the available literature more thoroughly as well as work with meta-reviews and meta-analyses to identify researchers' degrees of freedom in research on the privacy paradox. Preregistering conceptual decisions and analysis plans as well as sharing data, code and material can make these choices more transparent (cf. Dienlin et al., 2020). Understanding the impact of conceptual and analytical decisions on results and inferences will help future research to identify best practices for analyzing specific hypotheses and, in turn, advance true cumulative knowledge creation. For the time being, it seems likely that different conceptual and analytical choices may partly explain the inconsistencies in published studies investigating the privacy paradox.

Notes

1. Based on the search query 'privacy paradox' and a date range from 2004 to today on February 22nd, 2021.
2. I analyzed archival data that are not under my direct control, but can be downloaded via the GESIS Data Catalogue (European Commission, 2013, 2018, 2019a). Analysis scripts and an additional online supplement with all item formulations, descriptive analyses, and zero-order correlations can be assessed on the Open Science Framework: https://osf.io/m72gb/.
3. Additional analyses in which the different information types are treated as varying specifications of the dependent variable are presented in Figure A2 and A3 in the online supplement: https://osf.io/v85xf/.

Disclosure statement

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

Notes on contributor

Philipp K. Masur is an Assistant Professor for Persuasive Communication in the Department of Communication Science at the Vrije Universiteit Amsterdam. He earned his PhD in the social sciences at the University of Hohenheim. His research focuses on different aspects of computer-mediated communication. More specifically, he investigates social influence and persuasion processes on social media, privacy and self-disclosure in networked publics, and media and communication effects on individual well-being.

References

Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the facebook. In D. Hutchison et al (Ed.), *Privacy enhancing technologies* (pp. 36–58). Springer. doi:10.1007/11957454.

Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making. *IEEE Security and Privacy Magazine*, *3*(1), 26–33. https://doi.org/10.1109/MSP.2005.22

Altman, I. (1976). Privacy: A conceptual analysis. *Environment and Behavior*, *8*(1), 7–29. https://doi.org/10.1177/001391657600800102

Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: An empirical evaluation of informaiton transparency and the willingness to be profiled online for personalization. *Mis Quarterly*, *30*(1), 13–28. https://doi.org/10.2307/25148715

Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. *First Monday*, *11*(9). https://doi.org/10.5210/fm.v11i9.1394

Barth, S., & de Jong, M. D. T. (2017). The privacy paradox – investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and Informatics*, *34*(7), 1038–1058. https://doi.org/10.1016/j.tele.2017.04.013

Barth, S., de Jong, M. D. T., Junger, M., Hartel, P. H., & Roppelt, J. C. (2019). Putting the privacy paradox to the test: Online privacy and security behaviors among users with technical knowledge, privacy awareness, and financial resources. *Telematics and Informatics*, *41*, 55–69. https://doi.org/10.1016/j.tele.2019.03.003

Baruh, L., Secinti, E., & Cemalcilar, Z. (2017). Online privacy concerns and privacy management: A meta-analytical review. *Journal of Communication*, *67*(1), 26–53. https://doi.org/10.1111/jcom.12276

Bazarova, N. N., & Masur, P. K. (2020). Towards an integration of individualistic, Networked, and institutional approaches to online disclosure and privacy in a Networked ecology. *Current Opinion in Psychology*, *36*, 118–123. https://doi.org/10.1016/j.copsyc.2020.05.004

Blank, G., Bolsover, G., & Dubois, E.. (2014). *A New privacy paradox: Young people and privacy on social network sites.* Prepared for the Annual Meeting of the American Sociological Association, 17 August 2014, San Francisco, California. https://doi.org/10.2139/ssrn.2479938

Chen, H.-T. (2018). Revisiting the privacy paradox on social media with an extended privacy calculus model: The effect of privacy concerns, privacy self-efficacy, and social capital on privacy management. *American Behavioral Scientist*, *62*(10), 1392–1412. https://doi.org/10.1177/0002764218792691

Chen, H.-T., & Chen, W. (2015). Couldn't or wouldn't? The influence of privacy concerns and self-efficacy in privacy management on privacy protection. *Cyberpsychology, Behavior and Social Networking*, *18*(1), 13–19. https://doi.org/10.1089/cyber.2014.0456

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.

Del Giudice, M., & Gangestad, S. (2021). A traveler's guide to the multiverse: Promises, pitfalls, and a framework for the evaluation of analytic decisions. *Advances in Methods and Practices in Psychological Science*, *4*(1), 1–15. https://doi.org/10.1177/2515245920954925

De Wolf, R. (2020). Contextualizing how teens manage personal and interpersonal privacy on social media. *New Media & Society*, *22*(6), 1058–1075. https://doi.org/10.1177/1461444819876570

Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., Lukito, J., Bier, L. M., Zhang, R., Johnson, B. K., Huskey, R., Schneider, F. M., Breuer, J., Parry, D. A., Vermeulen, I., Fisher, J. T., Banks, J., Weber, R., Ellis, D. A., … de Vreese, C. (2020). An agenda for open science in communication. *Journal of Communication*, *71*(1), 1–26. https://doi.org/10.1093/joc/jqz052

Dienlin, T., Masur, P. K., & Trepte, S. (2019). *A longitudinal analysis of the privacy paradox.* https://doi.org/10.31235/osf.io/fm4h7

Dienlin, T., & Metzger, M. J. (2016). An extended privacy calculus model for SNSs—Analyzing self-disclosure and self-withdrawal in a U.S. representative sample. *Journal of Computer-Mediated Communication*, *21*(5), 368–383. https://doi.org/10.1111/jcc4.12163

Dienlin, T., & Trepte, S. (2015). Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors. *European Journal of Social Psychology*, *45*(3), 285–297. https://doi.org/10.1002/ejsp.2049

Dinev, T., & Hart, P.. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, *17*(1), 61–80. https://doi.org/10.1287/isre.1060.0080

Egelman, S., Felt, A. P., & Wagner, D. (2013). Choice architecture and smartphone privacy: There's a price for that. In R. Böhme (Ed.), *The economics of information security and privacy* (pp. 211–236). Springer. doi:10.1007/978-3-642-39498-0_10.

Epstein, D., & Quinn, K. (2020). Markers of online privacy marginalization: Empirical examination of socioeconomic disparities in social media privacy attitudes, literacy, and behavior. *Social Media+Society*, *6*(2), 205630512091685. https://doi.org/10.1177/2056305120916853

European Commission. (2011). *Special Eurobarometer 359: Attitudes on data protection and electronic identity in the European Union.* http://ec.europa.eu/public_opinion/archives/ebs/ebs_359_en.pdf

European Commission. (2013). *Eurobarometer 74.3 (2010)—Version 5.2.0 (2013)*. Gesis Data Catalogue DBK. doi:10.4232/1.11627.

European Commission. (2015). *Special Eurobarometer 431: Data protection*. http://ec.europa.eu/public_opinion/archives/ebs/ebs_431_en.pdf

European Commission. (2018). *Eurobarometer 83.1 (2015)—Version 2.0.0 (2018)*. Gesis Data Catalogue DBK. doi:10.4232/1.13071.

European Commission. (2019a). *Eurobarometer 91.2 (2019)—Version 1.0.0 (2019)*. Gesis Data Catalogue DBK. doi:10.4232/1.13318.

European Commission. (2019b). *Special Eurobarometer 487a: The General Data Protection Regulation*. https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/ResultDoc/download/DocumentKy/86886

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. Psychology Press.

Gambino, A., Kim, J., Sundar, S. S., Ge, J., & Rosson, M. B. (2016). User disbelief in privacy paradox: Heuristics that determine disclosure. In J. Kaye, A. Druin, C. Lampe, D. Morris, & J. P. Hourcade (Eds.), *CHI 2016* (pp. 2837–2843). The Association for Computing Machinery. doi:10.1145/2851581.2892413.

Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Hallam, C., & Zanella, G. (2017). Online self-disclosure: The privacy paradox explained as a temporally discounted balance between concerns and rewards. *Computers in Human Behavior*, *68*, 217–227. https://doi.org/10.1016/j.chb.2016.11.033

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, *64*, 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online social networks: Why we disclose. *Journal of Information Technology*, *25*(2), 109–125. https://doi.org/10.1057/jit.2010.6

Krasnova, H., Veltri, N. F., & Günther, O. (2012). Self-disclosure and privacy calculus on social networking sites: The role of culture. *Business & Information Systems Engineering*, *4*(3), 127–135. https://doi.org/10.1007/s12599-012-0216-6

Lutz, C., Hoffmann, C. P., & Ranzini, G. (2020). Data capitalism and the user: An exploration of privacy cynicism in Germany. *New Media & Society*, *22*(7), 1168–1187. https://doi.org/10.1177/1461444820912544

Lutz, C., & Strathoff, P. (2014). *Privacy concerns and online behavior – Not so paradoxical after all? Viewing the privacy paradox through different theoretical lenses*. Social Science Research Network. doi:10.2139/ssrn.2425132.

Madden, M., & Rainie, L. (2015). *Americans' attitudes about privacy, security and surveillance*. http://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/

Marwick, A. E., & boyd, d. (2014). Networked privacy: How teenagers negotiate context in social media. *New Media & Society*, *16*(7), 1051–1067. https://doi.org/10.1177/1461444814543995

Masur, P. K. (2018). *Situational privacy and self-disclosure: Communication processes in online environments*. Springer.

Masur, P. K. (2020). How online privacy literacy supports self-data protection and self-determination in the age of information. *Media and Communication*, *8*(2), 258–269. https://doi.org/10.17645/mac.v8i2.2855

Naughton, J. (2019, May 5). The privacy paradox: Why do people keep using tech firms that abuse their data? *The Guardian*. https://www.theguardian.com/commentisfree/2019/may/05/privacy-paradox-why-do-people-keep-using-tech-firms-data-facebook-scandal

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, *41*(1), 100–126. https://doi.org/10.1111/j.1745-6606.2006.00070.x

Nosko, A., Wood, E., & Molema, S. (2010). All about me: Disclosure in online social networking profiles: The case of faceboook. *Computers in Human Behavior*, *26*(3), 406–418. https://doi.org/10.1016/j.chb.2009.11.012

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, *3*(2), 173–182. https://doi.org/10.1038/s41562-018-0506-1

Petronio, S. (2002). *Boundaries of privacy*. State University of New York Press.

Reynolds, B., Venkatanathan, J., Gonçalves, J., & Kostakos, V. (2011). Sharing ephemeral information in online social networks: Privacy perceptions and behaviours. In P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, & M. Winckler (Eds.), *Human-Computer interaction – INTERACT 2011* (Vol. 6948, pp. 204–215). Springer.

Rohrer, J. M., Egloff, B., & Schmukle, S. C. (2017). Probing birth-order effects on narrow traits using specification-curve analysis. *Psychological Science*, *28*(12), 1821–1832. https://doi.org/10.1177/0956797617723726

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behavior*, *4*(11), 1208–1214. https://doi.org/10.1038/s41562-020-0912-z

Spiekermann, S., Grossklags, J., & Berendt, B. (2001). E-privacy in 2nd generation e-commerce: Privacy preferences versus actual behavior. *CACM*, *48*(3), 38–47. https://ssrn.com/abstract=761107

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Taddei, S., & Contena, B. (2013). Privacy, trust and control. Which relationships with online self-disclosure? *Computers in Human Behavior*, *29*(3), 821–826. https://doi.org/10.1016/j.chb.2012.11.022

Taddicken, M. (2014). The `privacy paradox' in the social web: The impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self-disclosure. *Journal of Computer-Mediated Communication*, *19*(2), 248–273. https://doi.org/10.1111/jcc4.12052

Trepte, S. (2020). The social media privacy model: Privacy and communication in the light of social media affordances. *Communication Theory. Advance Online Publication*, https://doi.org/10.1093/ct/qtz035

Trepte, S., Reinecke, L., Ellison, N. B., Quiring, O., Yao, M. Z., & Ziegele, M. (2017). A cross-cultural perspective on the privacy calculus. *Social Media+Society*, *3*(1), 1–13. https://doi.org/10.1177/2056305116688035

Tufekci, Z. (2008). Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology & Society*, *28*(1), 20–36. https://doi.org/10.1177/0270467607311484

Utz, S., & Krämer, N. (2009). The privacy paradox on social network sites revisited. The role of individual characteristics and group norms. *Journal of Psychosocial Research on Cyberspace*, *3*(2), Article 2. http://cyberpsychology.eu/view.php?cisloclanku=2009111001&article=2

Walrave, M., Vanwesenbeeck, I., & Heirman, W. (2012). Connecting and protecting? Comparing predictors of self-disclosure and privacy settings use between adolescents and adults. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *6*(1), Article 3. https://doi.org/10.5817/CP2012-1-3

Westin, A. F. (1967). *Privacy and freedom*. Atheneum.