THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Does (customer data) size matter? Generating valuable customer insights with less customer relationship risk

Citation for published version:
Plangger, K, Marder, B, Montecchi, M, Watson, R & Pitt, L 2023, 'Does (customer data) size matter? Generating valuable customer insights with less customer relationship risk', *Psychology and marketing*. https://doi.org/10.1002/mar.21866

Digital Object Identifier (DOI):
10.1002/mar.21866

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Psychology and marketing

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 30. Jul. 2023

Check for updates

Received: 18 August 2022 | Accepted: 27 June 2023

DOI: 10.1002/mar.21866

RESEARCH ARTICLE

Psychology & Marketing WILEY

Does (customer data) size matter? Generating valuable customer insights with less customer relationship risk

Kirk Plangger[1] | Ben Marder[2] | Matteo Montecchi[1] | Richard Watson[3] | Leyland Pitt[4]

[1]King's Business School, King's College London, London, UK

[2]Edinburgh Business School, University of Edinburgh, Edinburgh, UK

[3]Digital Frontier Partners, Melbourne, Australia

[4]Beedie School of Business, Simon Fraser University, Vancouver, BC, Canada

Correspondence

Kirk Plangger, King's Business School, King's College London, Bush House, 30 Aldwych, London WC2B 4BG, UK.
Email: kirk.plangger@kcl.ac.uk

Funding information

British Academy

Abstract

Customer surveillance is a pervasive marketing practice that involves the collection, usage, and storage of customers' data from transactions, loyalty programs, and social media. Customer data are valuable to firms in gaining or maintaining an edge over competitors by developing superior customer insights that may assist product or service innovations. However, customer surveillance practices also risk customer relationships by potentially activating privacy and data security concerns. This article explores customer insight strategies that focus customer surveillance by assessing the insight value of data sources to avoid unnecessary data collection and capture. Three prediction experiments show that three distinct data source attributes, namely *data quantity*, *data detail*, and *data content*, are diagnostic of the prediction accuracy of customer psychographic characteristics and behavioral intentions. By demonstrating that customer insights are more (or less) valuable when derived from different data sources, this article shows that "more" data is not necessarily better. We advocate a smarter approach to customer surveillance practices that are selective in choosing to capture customer data that can yield more accurate customer insights while reducing the risk of jeopardizing customer relationships.

KEYWORDS

customer data, customer insight value, customer privacy and data security concerns, customer relationships, customer surveillance, prediction experiments

1 | INTRODUCTION

Are bigger data better? Customer data have become currency, with many firms generally holding the belief that more quantities of data are better at developing more valuable customer insights (Bansal et al., 2016; Qi et al., 2016). This is reflected in the staggering growth of the global big data analytics industry that is projected to reach $650 billion by 2029 (Fortune, 2023). The widespread belief that bigger data can produce better insights has been exacerbated as the global proportion of customers shopping on digital platforms increased dramatically due to pandemic restrictions and is likely to grow to about one-third of all shoppers by 2025 (Colback, 2023), thus further simplifying customer data collection and capture. These data can produce customer insights that can improve performance by allowing firms to, for example, evaluate campaigns and channel effectiveness, design and test new products, and make demand forecasts (LaValle et al., 2011; McAfee & Brynjolfsson, 2012). However, despite efforts to collect and process more customer data,

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2023 The Authors. *Psychology & Marketing* published by Wiley Periodicals LLC.

less than half of data analytics managers believe that their team provides real value to their firm (Gartner, 2023).

These customer data result from the surveillance of customers' digital footprint, and the collection, usage, and storage of these data (Chandra et al., 2022; Lyon, 2007; Scarpi et al., 2022). Customer surveillance is increasingly less obtrusive, less costly, and more data-rich due to advances in technology, including, for example, facial emotion recognition scanners, location tracking devices, mobile commerce, biometric payment systems, social media platforms, and voice-controlled devices (e.g., Lim et al., 2022; Moriuchi, 2021, 2023). However, such surveillance activities have the potential to harm customer relationships by activating customers' privacy and security concerns (Lefkeli et al., 2022; Maseeh et al., 2021; Okazaki et al., 2020). For example, a majority of American consumers are concerned about social media platforms collecting personal data or tracking online behaviors, leading to over 90% of consumers reporting taking active measures to limit surveillance (Cusson, 2023).

Firms are motivated to increase customer surveillance, in essence, casting a broad net that gathers all data possible. However, like in the fishing industry, much of what is caught in the surveillance net may be of little or no value. Confronted with a large amount of data, managers face the complex task of discerning which data sources provide real value to their marketing operations (Clarke, 2016; Lefkeli et al., 2022). Furthermore, such a broad-net customer surveillance approach is likely to result in the activation of privacy concerns that damage customer relationships with the firm while producing limited insights.

To reduce relationship risk arising from privacy and security concerns, firms need to rethink and develop more efficient and effective market intelligence strategies (Plangger & Montecchi, 2020; Rowe et al., 2020). Some customer surveillance activities could be reduced or eliminated without adversely impacting the value of data sources by collecting more effective data from sources that better predict desired customer insights. In doing so, firms could increase the effectiveness of customer surveillance in providing customer insights that keep their products and services competitive while reducing the risk to customer relationships.

At present, many managers largely hold the notion that "bigger data" is better, yet are also aware of growing concerns related to customer privacy and the potential negative relationship effects (Clarke, 2016; Lefkeli et al., 2022). Addressing these surveillance trade-offs, our aim is to examine possibilities for smarter strategies for customer surveillance that optimize potential value against potential costs. Specifically, this article explores the extent to which different customer data sources predict customer insights to explicate how to partially resolve this paradox by reducing the need for broad customer surveillance. We use three experiments to assess the value of four customer data sources (credit statements, iTunes records, Facebook public profile, and Facebook detailed profile) to predict three customer insights: customer personality, high involvement purchase likelihood, and low involvement purchase likelihood. These customer data sources vary in their quantity (high, low), detail (high, low), and content (transaction, social media). Results confirm that customer insight prediction accuracy varies with the nature of the customer data source, and therefore the accurate selection of these sources may yield competitive advantages.

This article offers both conceptual and methodological contributions to the marketing and business literature. First, our conceptual development and results identify three attributes of data sources that have significant impacts on the accuracy of customer insight predictions. This suggests that firms should select the nature of their data sources depending on the type of insight they are seeking. Moreover, managers should audit current customer data that their firms possess to enable better customer insight predictions with fewer data sources. Second, our use of experimental methods to predict customer insights from different data sources demonstrates the unique utility of crowdsourcing to offer a "low-tech" method of gaining an understanding of complex consumer phenomena when human judgment is required.

2 | LITERATURE REVIEW

Firms use customer data to design, evaluate, promote, and refine their products and services to better meet customer needs. However, customers sometimes resent having their personal data collected or worry if their data is securely stored (Aboulnasr et al., 2022; Lyngdoh et al., 2023). This section explores why and how customer relationships are threatened by customer surveillance on digital platforms, discusses a method of structuring market intelligence, and examines different attributes of data sources.

2.1 | Customer surveillance and customer relationship risk

Customer surveillance is the gathering, analyzing, and storing of personal data through digital technologies about individuals, who may be aware or not aware of this data transfer (Park et al., 2015; Plangger & Montecchi, 2020). Such surveillance occurs to provide "value from the large untapped pools of data in the digital universe" to the surveillant (Gantz & Reinsel, 2012, p. 3). Marketing strategies based on intelligence gathered through customer surveillance generally outperform strategies based on managerial intuition or experience (LaValle et al., 2011; McAfee & Brynjolfsson, 2012). Market-oriented (i.e., focused on customer needs) firms are often known to perform better than product-oriented (i.e., focused on product capabilities) because of their ability to produce products and services that better meet customers' needs (Jaworski & Kohli, 1993). By generating, disseminating, and responding to customer insights derived from market intelligence, employee commitment to the firm (Jaworski & Kohli, 1993), customer satisfaction (Harter et al., 2002), customer loyalty (Salanova et al., 2005), and positive word-of-mouth (Filieri et al., 2018) have all been shown to increase. These positive outcomes of market intelligence that stem from customer surveillance enable firms to build long-term, intimate, and profitable customer relationships.

15206793, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mar.21866 by Edinburgh University, Wiley Online Library on [12/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Built on trust and commitment, customer relationships underpin customer loyalty and customer satisfaction (Morgan & Hunt, 1994). Customer trust refers to customers' confidence that a firm is reliable and has integrity. Customer commitment describes the importance of the relationship to customers and that they devote resources to maintain and perhaps enhance this relationship. If customers perceive a firm to be reliable and honest, in addition to the feeling that the relationship with the firm is important and valuable, an intimate, long-term customer relationship is likely to emerge.

On the one hand, as evidenced by the popularity of loyalty programs and the widespread acceptance of digital cookies, many customers frequently share their personal data with firms that they have favorable customer relationships with (Li et al., 2021; Plangger & Montecchi, 2020). However, if this surveillance leads to a data security event (e.g., hack, data device loss, unintended disclosure), it may activate customers' privacy concerns, or anxiety over the potential misuse or access of personal data (Bright et al., 2021; Lyngdoh et al., 2023). When activated due to a data security event, privacy concerns have the potential to dramatically impact customer relationships leading to an increase in switching behaviors (Martin et al., 2017; Rehman et al., 2020) and potentially spreading to other customers (Visentin et al., 2021).

On the other hand, firms often do not have strong customer relationships (Bauer, 2023), especially if they are new to an industry or have had data security issues in the past. If such a firm asks potential customers to disclose personal data, it may activate customers' privacy concerns, or anxiety regarding the ability to control when, how, and to what extent personal data is shared with others (Kumar et al., 2022; Malhotra et al., 2004; Smith et al., 1996). Activating privacy concerns can have negative consequences for trust in a firm, disclosure likelihood, attitudes toward the firm, heuristic and misattribution formation, social media engagement, and other behavioral intentions (Bright et al., 2021; Norberg & Horne, 2007; Rehman et al., 2020; Smith et al., 2011).

Firms require the gathering of intelligence to be market-orientated, which enables the production of enhanced products and services that better meet customers' needs. Successfully meeting the needs of customers allows the establishment or enhancement of customer relationships that are built on trust and commitment (McKechnie et al., 2018; Morgan & Hunt, 1994). If customers experience privacy threats from customer surveillance activities, customers' privacy and security concerns will be activated (Ioannou et al., 2021) leading to potentially detrimental effects on their relationship with the offending firm (Okazaki et al., 2009; Rehman et al., 2020), as well as the firm's reputation (Daudigeos et al., 2020). To reduce the risk of privacy threats, customer surveillance activities must be carefully and strategically conducted.

2.2 | Strategic market intelligence

Market intelligence has become a central aspect of marketing activities (e.g., customer relationship management systems, customer loyalty programs, user experience) (e.g., Plangger & Montecchi, 2020; Watson et al., 2023). However, many firms are collecting, capturing, storing, and using customers' personal data without well thought out market intelligence strategies that seek customer data for specific customer insights (Plangger & Watson, 2015; Plangger et al., 2022; Qi et al., 2016; Turow, 2008). This situation is exacerbated by technological advances that have made customer surveillance more powerful, less visible, and less expensive (Bauman & Lyon, 2013).

Market intelligence's data sources can be more efficiently gathered by collecting customer data according to surveillance prompts. These prompts categorize discrete customer facts collected by customer surveillance activities using a set of generic questions or prompts (i.e., when, where, what, how, who, why, and outcome; Plangger & Watson, 2015; Thomsen, 2002). "When" uncovers the temporal nature of customer behavior by understanding the frequency, time, or date of customer activity. "Where" holds the physical or virtual locations of customers. "What" is essential for firms to manage inventory stocks, and to determine which offerings are frequently bought (or not bought) together. "How" aids the understanding of customers' preferred methods of customer activity, including shopping orientations, payment type choices, and other potential customer (dis)satisfaction points. "Who" can be used to create unique customer profiles that might include characteristics such as interests, demographics, psychographics, memberships, and links to other customers. Together the surveillance prompts provide a picture of customers from a variety of different perspectives that can yield many valuable customer insights.

Routine transaction data (e.g., point of purchase records, electronic receipts) can satisfy many of the surveillance prompts. To add further depth and aid understanding, transaction data could be augmented with sensor data (e.g., face recognition, RFID tags) to extract additional and perhaps more precise insights. However, even with additional sensor data, transaction data provide a limited understanding of customers' motivations (i.e., why), which are most often captured through more traditional means (e.g., surveys, focus groups). While useful in many instances, such methods may carry measurement issues that bias findings (e.g., social desirability bias, expectation bias; Creswell & Creswell, 2017). Social media offer a potential solution, as they may provide clues to customers' motivations through the surveillance of customer forums, check-in to locations, and profile histories (Kietzmann et al., 2011; Marder, 2018). These social media data sources can be mined to identify possible motivators and married with other data sources to potentially give a more detailed picture of customers and their behaviors (Micu et al., 2017; Qi et al., 2016).

Firms may not be able to anticipate all the potential customer insights that they may need to seek when collecting customer data. By designing marketing intelligence using surveillance prompts and strategically selecting appropriate customer data sources that satisfy these prompts, firms will be able to develop unimagined customer insights (Ghanbarpour et al., 2022; Plangger & Watson, 2015; Schweidel et al., 2022; Watson, 2013). In doing so, firms can examine their customer surveillance activities and reduce those activities that do not add more information to satisfy surveillance prompts.

15206793, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mar.21866 by Edinburgh University, Wiley Online Library on [12/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.3 | Customer insights and data sources

Since there is a potential risk to customer relationships, firms must be strategic in the design of market intelligence by seeking customer data sources that provide the most valuable customer insights. Customer data come from a variety of sources, including, for example, transaction records, CCTV observations, loyalty programs, advanced sensors, and social media interactions (Plangger & Watson, 2015). *Customer insights* are identified patterns in customer data that indicate customers' personalities, future purchases, preferences, needs, and other customer attributes (Chandra et al., 2022; Kohli & Jaworski, 1990).

Customer insights are valued by their usefulness within a particular context, and insight usefulness increases with accuracy (Hess & Doe, 2013). The constructivist approach to accuracy has been widely used in past research (Hall et al., 2007), and it examines the level of agreement or consistency among individual judges (Funder, 1995; Kruglanski, 1989). However, since many judges can agree on a prediction and yet still be incorrect, prediction accuracy is a more valuable indicator of customer insight value. Prediction accuracy can be measured using trusted comparison values from other sources of data that are deemed to be (more) accurate. The more accurate the predicted customer insight, the more valuable that insight is to firms seeking to better target a customer group.

Both through technology and human resources, firms can predict a range of customer insights, including personality characteristics (i.e., who) and future purchases (i.e., what) among others, to improve the effectiveness and efficiency of advertising and marketing strategies (Campbell et al., 2022; Hess & Doe, 2013; Trivedi & Teichert, 2021). For example, analysis of social media posts assisted by machine learning algorithms can predict self-monitoring characteristics (He et al., 2014). These predicted customer insights are based on human or algorithmic judges' attributions or links between observations and casual explanations made from examining the available data (Folkes, 1988). Customer data sources are heterogeneous in nature, and, intuitively, some data sources predict specific customer insights better than others. This article explores the appropriateness of digital data sources to accurately predict customer insights by examining the effects of three data source attributes, namely data quantity, data detail, and data content.

Data quantity and detail are important data source attributes to assess the knowledge contained in a set of data. If judges have more knowledge about the subject of the data, empirical evidence shows that those judges make more accurate predictions of customer insights than judges with less knowledge (Funder, 1995). *Data quantity* refers to the sheer amount of data points in a data set. *Data detail* involves the specificity or granularity of the data points in a data set. For example, credit card statements contain relatively less detail about a transaction than a store's transaction records, yet if a credit card is frequently used, it may contain a higher quantity of data than the transaction record of only a single store. Customer data sources that are high in both quantity and detail contain more potential knowledge about customers, and thus may provide more accurate and consistent predictions of some customer insights. *Data content* describes the objective subject matter or substance of a data set.

Although there are other customer data sources, this article builds on recent work (e.g., Bansal et al., 2016; Dimitriu & Guesalaga, 2017; Ghanbarpour et al., 2022; Pitt et al., 2020) on developing customer insights from customer data. Specifically, it examines social media and transaction data in the context of predicting personality characteristics or purchase likelihood in high or low involvement contexts. Individuals predict personality characteristics very quickly after first meeting a new person, even without pre-existing knowledge about that person. These predictions are often fairly consistent and accurate impressions of that person's personality (Uleman, 1999), thus data sources that contain information about an individual's nature should provide more accurate personality predictions. Purchase likelihood predictions require different knowledge about the past behavior of that person, as data sources of past purchase behavior have been shown to improve to future purchase behavior predictions (Ajzen, 2011).

2.4 | An exploratory framework of data source attributes

Firms need customer insights to enhance their business efficiency and effectiveness by better understanding target customers (Hess & Doe, 2013; Kohli & Jaworski, 1990). Customer surveillance on digital platforms is a ubiquitous phenomenon that gives rise to a large amount of data that is now shaping the future of such intelligence gathering (Gantz & Reinsel, 2012). However, developing enhanced surveillance capabilities risks damaging relationships with customers by violating their privacy (Martin et al., 2017; Maseeh et al., 2021; Plangger & Montecchi, 2020; Okazaki et al., 2020). To address such risk, firms may potentially utilize surveillance prompts and be critical of the value of customer insights derived from customer data sources to reduce the need for extensive customer surveillance.

At present little empirical research has explored these strategic tradeoffs, with the aim of providing firms with a better understanding of how to optimize the value gained from customer surveillance against the potential privacy costs of doing so. Addressing this deficiency, this article examines how three critical data source attributes (data quantity, data detail, and data content) can be used to differentiate source value in terms of customer insight prediction accuracy. Our exploratory framework (Figure 1) guides the development of three empirical studies that examine the effect of different configurations of data source attributes on the prediction accuracy of psychographic (personality traits) and behavioral intention (purchase likelihood) customer insights.

3 | EMPIRICAL STUDIES

This section reports the method and results of three experiments that expose respondents to random sets of customer data and ask them to predict either personality (Study 1) or purchase likelihoods

15206793, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mar.21866 by Edinburgh University, Wiley Online Library on [12/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

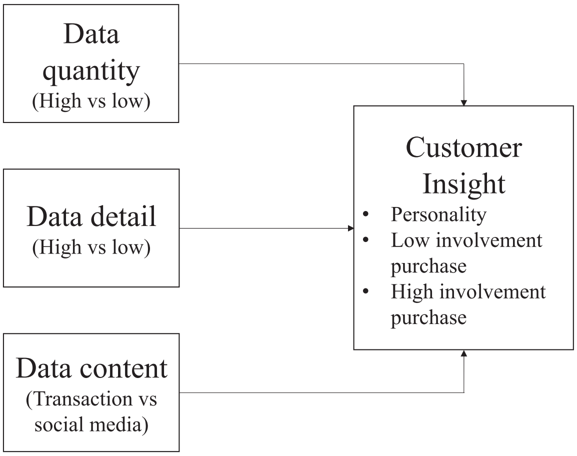

FIGURE 1 Exploratory framework.

TABLE 1 Data sources' attributes.

Data source	Data quantity	Data detail	Data content
Credit card	High	Low	Transaction
iTunes	Low	High	Transaction
Facebook public	Low	Low	Social media
Facebook detail	High	High	Social media

(Study 2 and Study 3). These customer insight predictions are then averaged and compared to predictions from other sources to assess the value of different data sources. In doing so, these experiments illustrate a method of evaluating data sources by determining their value in predicting customer insights. As human respondents are used in the place of algorithms, artificial intelligence, or computers, this method is a return to manual analysis and interpretation that underpins turning data into knowledge (Fayyad et al., 1996).

These experiments used the same four data sources from four sets of an individual's personal data: (1) one month of credit card statements, (2) three months of iTunes purchase records, (3) public Facebook data using minimal privacy settings, and (4) detailed Facebook data downloaded from Facebook account settings. In each study, respondents were randomly allocated to one of four conditions representing the different personal data sources. These data sources varied in terms of all three data attributes (i.e., data quantity, data detail, and data content) discussed above (see Table 1). Data quantity and data detail are relative attributes and are only useful in comparison with other data that are either higher or lower in that attribute. Data content is an objective attribute that describes the source of the data as either from social media or transaction records.

Relative customer insight value can be measured by comparing the average prediction accuracy of data sources. Using these measures, these studies explore the contribution of data source attributes to customer insight value. To assess prediction accuracy, comparison scores were collected from the individual who provided the data (self-reported) and eight close friends and family members. Comparisons of predicted scores and friends and family scores evaluate the accuracy of personality predictions, as close acquaintances' personality predictions are generally more valid than self-reported predictions (Kolar et al., 1996). As they did not access any of the data sources provided to experimental respondents, friends and family assessed personality from experiences and interactions with the individual who provided the data for this research. In contrast, purchase likelihood predictions were evaluated using self-reported scores, because of the additional personal knowledge about purchase intentions and firm preferences that the individual has access to make predictions.

Each data source's relative customer insight value was assessed by comparing data sources' prediction accuracy scores. Specifically, average prediction values from each data source were compared to the corresponding self-reported or friends and family score depending on the study. The absolute value of the differences between the predicted and comparison scores of each scale item were summed to provide a measure of how inaccurate the predictions are from the comparison values. Then, analysis of variance (ANOVA) tests on the inaccuracy measure indicated the level of variation between conditions. Planned inaccuracy mean contrasts tested the significance of data source attributes (high vs. low data detail/quantity, transaction vs. social media data content) contribution to prediction accuracy (see Table 1). We calculated and reported effect sizes for all significant data attributes using Cohen's *d* statistic. The specific customer insights or scores are of little value in the context of the studies but would be of great value to a firm or perhaps in other research contexts.

3.1 | Study 1: Predicting customer personality

3.1.1 | Materials

Study 1 asked respondents to observe an individual's personal data and assess that individual's personality using the Gilbert and Warren (1995) personality segmentation scale that includes five dimensions: economizer, credit user, self-confidence, home-oriented, and fashionable (see Appendix A). We choose this scale because of its simplicity and the range of identified characteristic dimensions, as well as the more actionable managerial implications for consumer segmentation compared to a general psychology scale (e.g., the Big Five). Respondents' personality predictions were evaluated for accuracy using two sets of comparison scores (self-reported and friends and family scores; see Table 2). The actual personality prediction scores are not relevant for this study, as the purpose is to compare the accuracy of scores between data sources.

3.1.2 | Manipulation and attention checks

All respondents underwent several tests to both check experimental manipulations and respondents' attention to ensure response quality.

15206793, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mar.21866 by Edinburgh University, Wiley Online Library on [12/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Prediction inaccuracy results.

Data source	Study 1 *n*	Self	F&F	Study 2 *n*	Self	Study 3 *n*	Self
Credit card	38	6.23	4.89	30	3.93	34	4.03
iTunes	33	5.61	4.31	33	5.55	38	5.45
Facebook public	37	5.60	3.88	36	5.11	38	3.68
Facebook detail	40	5.01	3.66	29	4.55	37	5.05
Total	148			128		147	

Note: Self, self-reported scores; F&F, friends and family scores.

They were asked to identify the kind of data they had observed (e.g., Facebook, credit card statements, iTunes records, or other). Furthermore, respondents were asked to select "agree" in a question scaled from "strongly disagree" to "strongly agree" to check attention near the end of the survey. Respondents that failed these checks were removed from the sample. Attempts were made to recruit at least 40 respondents for each of the data source conditions to ensure sufficient statistical power after the removal of incomplete responses and failed manipulation and attention checks.

3.1.3 | Sample

Study 1 instructed respondents to evaluate an individual's data to make specific predictions about that individual. As any English-speaking adult could be a potential respondent, all respondents in this and the following experiments were recruited from an online consumer panel pool using the *Cloud Research* service, commonly employed in research in the marketing and business literature (Erz et al., 2018; Hulland & Miller, 2018; Kees et al., 2017), with the only restrictions being that they lived in North America and were over 19 years of age. While not perfectly representative of the North American population, empirical evidence shows that samples using similar online services are not dramatically skewed or biased in comparison with other online and offline survey collection methods (Goodman et al., 2013; Kees et al., 2017).

Initially, 185 survey responses were collected, and after cleaning the data set of incomplete responses, nonunique IP addresses, and failed manipulation or attention checks, the resulting cleaned data set contained 121 responses. This resulted in a usable response rate of 65.4% over all of the data sources, and no systematic bias was apparent in deleted responses. The sample was 57.9% female, 62.0% under 40 years of age, 50.9% single, and 75.3% of European descent.

3.1.4 | Procedure

Respondents were asked to answer a survey that was laid out using the following procedure: (1) accept the informed consent form; (2) observe one of four randomly allocated sets of an individual's personal data; (3) answer questions to predict the personality of an individual based on their observations of their data; and (4) answer demographics questions. Respondents received a nominal incentive ($0.60 on average) for their participation to obtain an adequate number of responses.

3.1.5 | Results

Results of a one-way ANOVA test indicate significant differences in prediction accuracy among data sources when evaluated against self-reported (F (3, 121) = 2.994, p = 0.034) and Family and Friends (F (3, 121) = 4.640, p = 0.004) comparison values. The personality predictions are arithmetically closer to the friends and family comparison scores than self-reported values. This finding is in line with the literature (see Kolar et al., 1996) and indicates the difficulty of individuals to objectively assess their own personality. Therefore, only inaccuracy statistics that use the friends and family comparison scores are used for further statistical analysis. Planned inaccuracy mean contrasts reveal that social media data content had a significant impact on prediction inaccuracy (social media vs. transaction: t (121) = −3.324, p = 0.001, d = −0.603), but data quantity (high vs. low: t (121) = 0.730, p = 0.467), and data detail (high vs. low: t (121) = −1.574, p = 0.112) did not.

3.2 | Pretest: Selecting brands for involvement purchase likelihood studies

The brands used as stimuli in the purchase behavior prediction studies (Studies 2 and 3) need to be dissimilar enough to allow for potential variation. Thus, we surveyed 116 respondents to rate brands on perceived value attributes (value for money, functional performance, good service, social status, value expression, and reputation; Sweeney & Soutar, 2001). Also, pretest respondents were asked to assess the likelihood that a consumer would buy brands in combination with one another. From these results, we selected three low-involvement brands (Starbucks coffee, Red Bull energy drink, and Miller Lite beer) and three high-involvement brands (United Airlines business class service, Mercedes, and Apple iPhone). These brands were used in Study 2 (low involvement) and Study 3 (high involvement).

3.3 | Study 2: Predicting low involvement purchase likelihood

3.3.1 | Materials

Study 2 asked respondents to observe an individual's personal data to assess the purchase likelihood of the pretested low-involvement brands (i.e., Starbucks coffee, Red Bull energy drink, and Miller Lite beer). In line with Study 1's procedure, respondents were randomly placed into one data source condition by the survey software.

15206793, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mar.21866 by Edinburgh University, Wiley Online Library on [12/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

To ensure that respondents were not biased against or for the specific brands in the experiment, attitudes toward these brands were evaluated before observing the data sources using Homer's (1995) brand attitude scale (see Appendix A). The study assessed accuracy by comparing self-reported scores and predicted scores (see Table 2). Following Study 1's procedure, manipulation and attention checks were applied to ensure data quality.

3.3.2 | Sample

Initially, from *Cloud Research*, 157 responses were collected before removing incomplete responses, failed manipulation checks, failed attention checks, and nonunique responses. These removed responses had no apparent data source, date, or other systematic bias. This resulted in a cleaned data set containing 124 unique responses and a usable response rate of 80.0%. Respondents were predominantly male (56.0%), under 40 years of age (73.5%), university-educated (79.3%), and half were married (50.0%).

3.3.3 | Procedure

Respondents followed a similar procedure to Study 1: (1) accept the informed consent form; (2) observe one of four randomly allocated sets of an individual's personal data; (3) answer questions to predict purchase likelihoods of low-involvement brands based on their observations; and (4) answer demographics questions. Respondents received a nominal incentive ($0.60 on average) for their participation.

3.3.4 | Results

The mean inaccuracy of purchase predictions varies significantly across data sources, as tested by one-way ANOVA (F (3, 124) = 7.036, $p < 0.001$). Planned inaccuracy mean contrasts reveal that data quantity (high vs. low: t (124) = −4.142, $p < 0.001$, $d = -0.749$) and data detail (high vs. low: t (124) = 2.008, $p = 0.047$, $d = 0.356$) are significant data attributes when predicting low involvement purchase likelihood. Data content (social media vs. transaction: t (124) = 0.351, $p = 0.728$) is not a significant data attribute.

3.4 | Study 3: Predicting high involvement purchase likelihood

3.4.1 | Materials

Study 3 asks respondents to predict purchase likelihood of the three pretested high purchase involvement brands (i.e., United Airlines business class service, Mercedes automobiles, and Apple iPhones) by randomly observing one of four data sources from an individual. Like Study 2, respondents' brand attitudes were measured using Homer's (1995) scale to guard against systematic bias within the data source conditions. This study assessed accuracy by comparing self-reported scores and predicted scores (see Table 2). Following Study 1's and 2's procedures, manipulation and attention checks were applied to ensure data quality.

3.4.2 | Sample

Initially, from *Cloud Research*, 165 responses were collected before the data was cleaned to remove incomplete responses, failed manipulation checks, and nonunique responses. These removed responses had no apparent data source, date, or other systematic bias. This resulted in a data set containing 143 unique responses and a usable response rate of 86.7%. The sample respondents were predominantly female (60.7%), under 40 years of age (62.2%), university educated (62.3%), of European descent (73.4%), and 42.3% were married.

3.4.3 | Procedure

Similar to Studies 1 and 2, respondents followed this procedure: (1) accept the informed consent form; (2) observe one of four sets of an individual's personal data; (3) answer questions to predict purchase likelihoods of high involvement brands based on their observations; and (4) answer demographics questions. Respondents received a nominal incentive ($0.60 on average) for their participation.

3.4.4 | Results

Respondents' predictions vary significantly in terms of average inaccuracy means between data sources tested by one-way ANOVA (F (3, 143) = 4.707, $p = 0.004$). Planned inaccuracy mean contrasts reveal that data detail (high vs. low: t (143) = 3.599, $p < 0.001$, $d = 0.594$) is a significant factor, and data quantity (high vs. low: t (143) = −0.062, $p = 0.951$), and content (social media vs. transaction: t (143) = −0.954, $p = 0.342$) are not.

4 | GENERAL DISCUSSION

This section compares and contrasts the results of the three experiments to reflect on the contribution of the data source attributes. The following paragraphs assess the data source attributes for their contribution to customer insight value using Cohen's d (see Table 3) that measures effects in terms of standard deviations. For example, a $d = 0.1$ effect means that the effect of the treatment or data source is 0.1 standard deviation. Cohen (1992) offers a simple scale to describe the size of effects: small effect $d = 0.2$, moderate effect $d = 0.5$, and large effect $d = 0.80$.

15206793, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mar.21866 by Edinburgh University, Wiley Online Library on [12/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Effect sizes of prediction inaccuracy data source attributes.[a]

Customer insight	Study tested	Data source attributes High quantity	High detail	Social media content
Personality	1	NS	NS	−0.603
Low purchase involvement	2	−0.749	0.356	NS
High purchase involvement	3	NS	0.594	NS

[a]Effect sizes reported are calculated using Cohen's *d* that accounts for different sample sizes.

Data quantity has a moderate effect on low-involvement purchase likelihood prediction accuracy (Study 2). High-quantity customer data sources (credit card statements and detailed Facebook data) significantly increase the accuracy of low involvement purchase likelihood predictions by 0.749 of a standard deviation compared to low-quantity data sources (iTunes purchase records and public Facebook data). Study 1 shows that data quantity is not a significant factor in the accuracy of high involvement purchase likelihood or personality predictions.

Data detail has a significant moderate effect on purchase likelihood prediction accuracy in low (Study 2) and high (Study 3) involvement contexts. High-detail customer data sources (e.g., iTunes purchase records and detailed Facebook data) significantly decrease purchase behavior prediction accuracy by 0.356 (Study 2) and 0.594 (Study 3) standard deviations compared to low-detail data sources (credit card statements and public Facebook data). This is a surprising result, as the detail of the data should theoretically provide more information to respondents that should, in turn, improve prediction accuracy. Data detail did not significantly contribute to personality prediction accuracy (Study 1).

Data content has a significant moderate effect on customer personality prediction accuracy (Study 1). Social media data content (public and detailed Facebook data) significantly increases customer personality prediction accuracy compared to transaction data content (credit card statements and iTunes purchase records). Data source content does not significantly contribute to purchase likelihood predictions in both low- and high-involvement contexts.

In short, these results indicate that the value of a customer data source to predict a specific customer insight depends greatly on that insight. Thus, the selection of customer data sources is a very complex process and needs to be informed first by which customer insights are required using the surveillance prompt framework described above.

4.1 | Research contributions

This article offers both conceptual and methodological contributions. First, at the conceptual level, we theorize and empirically validate an exploratory framework that delineates three distinct data source attributes, namely data quantity, data detail, and data content. These attributes differentiate data sources in terms of prediction accuracy of important psychographic (i.e., personality) and behavioral intention (i.e., purchase likelihood) customer insights. We provide evidence that a targeted approach to intelligence gathering that accounts for the specific attributes of the data collected in relation to the prediction target yields more optimal results (Chandra et al., 2022; Plangger & Watson, 2015). Our findings push the discussion beyond data mining and big data (LaValle et al., 2011; McAfee & Brynjolfsson, 2012; Wedel & Kannan, 2016), and into the cultivation of actionable customer insights that can have a real impact on business strategy, thus showing that "bigger data" are not necessarily a better strategy. By introducing three critical data source attributes, we contribute to a growing stream of research (e.g., Bleier et al., 2020; Scarpi et al., 2022) that advocates for a more balanced data-driven approach to marketing strategy that limits the negative consequences of consumer privacy concerns (Bright et al., 2021; Ioannou et al., 2021; Maseeh et al., 2021).

Second, at the methodological level, we illustrate an experimental sequence that leverages crowdsourced responses to predict customer insights accuracy from different data sources. By applying crowdsourcing techniques (e.g., Conley & Tosti-Kharas, 2014; Gelper et al., 2018), this method offers a valid alternative to overcome time and other resource constraints when researchers want to employ human judgment in their data analysis workflows.

4.2 | Practical implications

Firms need to carefully consider how they generate customer insights, gather market intelligence, and conduct customer surveillance to reduce the risk to customer relationships. Furthermore, providers of the technologies used to obtain intelligence (e.g., Facebook, TikTok), must consider the management of their user data to strike a balance between optimizing value to third parties and safeguarding their users. Using surveillance prompts to structure market intelligence may reduce the need for extensive and obtrusive customer surveillance by selecting missing customer insights. Furthermore, evidence from three studies supports that customer data sources can be valued in terms of how accurately they predict customer insight. In what follows, we discuss the practical implications of our research for firms aiming to leverage a smarter customer surveillance approach to generate meaningful customer insights.

Data source content is a significant factor that has a moderate effect on personality prediction accuracy. In terms of customer personality predictions, Facebook data provide more accurate predictions than transaction data. Firms could apply this finding, for example, by micro-targeting customers to specifically appeal to personality groups (e.g., the product's value for money could be highlighted for customers that are high economizers in advertising), or by tailoring customer services to meet specific personality traits (e.g., providing additional remote or home-visit services for

15206793, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mar.21866 by Edinburgh University, Wiley Online Library on [12/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

customers that are highly home-oriented). Social media data have immense potential value for firms to discover many customer insights that have been difficult to predict using only transaction data (e.g., customer personality, purchase motivation, and firm usage) or self-reported survey data (Li et al., 2021; Marder, 2018). However, social media data sources are often unstructured in contrast to highly structured transaction or survey data, thus making deep analysis and interpretation more difficult (Pitt et al., 2019). Until technology advances, firms can utilize crowdsourced human intelligence to easily and cheaply process large social media data sets, as employed in the three studies presented in this article. Moreover, while social media data are often public and easily accessible, firms need to be aware of ethical considerations as they capture these data to reduce the potential negative impact on customer relationships (Boyd & Crawford, 2012).

Conversely, results from Study 2 indicate that data quantity is a significant contributor to purchase likelihood prediction accuracy for low-involvement brands. Thus, these brands may want to invest in customer data sources that are not necessarily detailed, but that capture a high quantity of data, such as credit card data, to predict purchase likelihood. As the results for personality and purchase likelihood are different, firms could calibrate customer surveillance activities by evaluating customer data sources using experiments to understand the prediction value of a specific customer insight. In doing so, firms can trial a customer data source to see its value before investing resources into it and potentially increasing the customer relationship risk.

4.3 | Limitations and future research directions

Our research offers a deep exploration of the customer insight value of three personal data sources (credit card statements, iTunes transaction records, Facebook data) from one individual. Researchers can use our findings and test customer insight value in other contexts with other kinds of individuals or alternative personal data sources (e.g., more detailed retail transaction data, loyalty program records, other types of social media data). The data sources used in the experiments spanned relatively short time periods (credit card: 1 month; iTunes: 3 months; Facebook: static screenshot printout). Thus, future studies could use panel customer data that may provide more accurate predictions of customer insights over time.

The experiments are also limited to predicting two forms of customer insights: customer personality (who) and purchase likelihood (what). Future research could examine other customer insight predictions using the same prediction experimental method, such as, for example, purchase motivation (why), location (where), preferred payment methods (how), purchase frequency (when), or willingness to pay (outcome). The results of these studies might provide much value for firms to better use market intelligence sources.

As with all crowdsourcing research applications, the analysis is limited to the mental capacity of individual respondents (Conley & Tosti-Kharas, 2014). Future research could examine what respondent attributes promote more accurate predictions. Furthermore, advanced algorithms or artificial intelligence could be employed in the future to recreate these experiments and remove the human dimension from customer insight prediction and valuation.

To further inform the selection of customer data sources, additional research is needed to examine how customers respond to various kinds of customer surveillance activities, as some types of customer data might be seen as more sensitive than others. For example, some customers may perceive public social media data as more private than transaction data collected at the point of purchase. Thus, understanding customer surveillance attitudes and preferences is important to better satisfying customer needs and creating even stronger customer relationships (Potoglou et al., 2017).

Firms need to understand and evaluate the ethical, reputational, customer relationship, legal, and other risks that underscore customer surveillance (Bonina et al., 2021). The findings described above point to the benefit of collecting and analyzing public social media data to predict customer personality, but firms need to understand the ethical implications and risks before conducting social media surveillance. Future research into the ethics of customer surveillance may reveal important implications for marketing, advertising, information systems, and management practice, as well as public policy.

4.4 | Concluding thoughts

Firms require intelligence resources to remain competitive because they enable innovation and improvement of products and services in line with customers' needs. However, if customers perceive a personal privacy threat due to surveillance activities, firms risk harming customer relationships. Thus, firms need to temper their desire for an abundance of customer data and carefully consider the efficiency and effectiveness of customer surveillance activities. Throughout this article, the problem of risking customer relationships to gain customer data has been highlighted again and again. Yet, many firms engage in unfocused customer surveillance by collecting and capturing large sets of customer data that make up a firm's market intelligence. In response, seeking to make firms' surveillance activities smarter, this article calls on surveillance prompts (who, what, where, when, why, how, and outcome) to structure market intelligence so it maintains its value potentially with fewer customer data sources. Then, the article develops a method of assessing customer data sources in terms of the accuracy of predicted customer insights to inform the selection of more effective customer data sources.

Through a series of three experiments, this article empirically explores the effectiveness of four different customer data sources in predicting customer personality and purchase likelihoods in low- and high-involvement brand contexts. The findings include the benefit of social media (Facebook) data in more accurately predicting customer personality, and high quantity (credit card or detailed Facebook data)

15206793, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mar.21866 by Edinburgh University, Wiley Online Library on [12/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

data in more accurately predicting low involvement brand purchase likelihood.

In conclusion, "bigger data" are not necessarily better. Firms can conduct efficient and effective customer surveillance by applying surveillance prompts to their market intelligence and evaluating potential customer data sources on the value of their predicted customer insights. The resulting market intelligence strategy enables product and service innovations while being sensitive to customer privacy and security concerns, thus reducing customer relationships risk.

ACKNOWLEDGMENTS

The authors would like to thank the British Academy for partly funding this research through their Small Research Grants scheme. Moreover, the authors would like to thank the King's Business School for providing initial seed corn funding for this research.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Kirk Plangger http://orcid.org/0000-0002-0354-9707
Matteo Montecchi http://orcid.org/0000-0001-5884-4503
Richard Watson http://orcid.org/0000-0003-0664-8337
Leyland Pitt http://orcid.org/0000-0002-3099-9164

REFERENCES

Aboulnasr, K., Tran, G. A., & Park, T. (2022). Personal information disclosure on social networking sites. *Psychology & Marketing*, *39*(2), 294–308.

Ajzen, I. (2011). The theory of planned behaviour: Reactions and reflections. *Psychology & Health*, *26*(9), 1113–1127.

Bansal, G., Zahedi, F. M., & Gefen, D. (2016). Do context and personality matter? Trust and privacy concerns in disclosing private information online. *Information & Management*, *53*(1), 1–21.

Bauer, B. C. (2023). Strong versus weak consumer-brand relationships: Matching psychological sense of brand community and type of advertising appeal. *Psychology & Marketing*, *40*(4), 791–810.

Bauman, Z., & Lyon, D. (2013). *Liquid surveillance: A conversation*. John Wiley & Sons.

Bleier, A., Goldfarb, A., & Tucker, C. (2020). Consumer privacy and the future of data-based innovation and marketing. *International Journal of Research in Marketing*, *37*(3), 466–480.

Bonina, C., Koskinen, K., Eaton, B., & Gawer, A. (2021). Digital platforms for development: Foundations and research agenda. *Information Systems Journal*, *31*, 869–902.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679.

Bright, L. F., Lim, H. S., & Logan, K. (2021). "Should I Post or Ghost?": Examining how privacy concerns impact social media engagement in US consumers. *Psychology & Marketing*, *38*(10), 1712–1722.

Campbell, C., Sands, S., Montecchi, M., & Jensen Schau, H. (2022). That's so instagrammable! understanding how environments generate indirect advertising by cueing consumer-generated content. *Journal of Advertising*, *51*(4), 411–429.

Chandra, S., Verma, S., Lim, W. M., Kumar, S., & Donthu, N. (2022). Personalization in personalized marketing: Trends and ways forward. *Psychology & Marketing*, *39*(8), 1529–1562.

Clarke, R. (2016). Big data, big risks. *Information Systems Journal*, *26*(1), 77–90.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Colback, L. (2023). The rise of the platform economy. *The Financial Times*. https://www.ft.com/content/e5f5e5b9-3aec-439a-b917-7267a08d320f

Conley, C., & Tosti-Kharas, J. (2014). Crowdsourcing content analysis for managerial research. *Management Decision*, *52*(4), 675–688.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Cusson, N (2023). *Forbes advisor*. The Fear of Internet Surveillance. https://www.forbes.com/advisor/business/fear-internet-surveillance/

Daudigeos, T., Roulet, T., & Valiorgue, B. (2020). How scandals act as catalysts of fringe stakeholders' contentious actions against multinational corporations. *Business & Society*, *59*(3), 387–418.

Dimitriu, R., & Guesalaga, R. (2017). Consumers' social media brand behaviors: Uncovering underlying motivators and deriving meaningful consumer segments. *Psychology & Marketing*, *34*(5), 580–592.

Erz, A., Marder, B., & Osadchaya, E. (2018). Hashtags: Motivational drivers, their use, and differences between influencers and followers. *Computers in Human Behavior*, *89*, 48–60.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*(3), 37–54.

Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information & Management*, *55*(8), 956–970.

Folkes, V. S. (1988). Recent attribution research in consumer behavior: A review and new directions. *Journal of Consumer Research*, *14*(4), 548–565.

Fortune. (2023). *Big data analytics market size, share, growth statistics*. Fortune Business Insights. https://www.fortunebusinessinsights.com/big-data-analytics-market-106179

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*(4), 652–670.

Gantz, J., & Reinsel, D. (2012). *The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east* (pp. 1–16). IDC iView: IDC Analyze the future.

Gartner. (2023). *Gartner survey reveals less than half of data and analytics teams effectively provide value to the organization*. Gartner. https://www.gartner.com/en/newsroom/press-releases/03-21-2023-gartner-survey-reveals-less-than-half-of-data-and-analytics-teams-effectively-provide-value-to-the-organization

Gelper, S., Peres, R., & Eliashberg, J. (2018). Talk bursts: The role of spikes in prerelease word-of-mouth dynamics. *Journal of Marketing Research*, *55*(6), 801–817.

Ghanbarpour, T., Sahabeh, E., & Gustafsson, A. (2022). Consumer response to online behavioral advertising in a social media context: The role of perceived ad complicity. *Psychology & Marketing*, *39*(10), 1853–1870.

Gilbert, F. W., & Warren, W. E. (1995). Psychographic constructs and demographic segments. *Psychology & Marketing*, *12*(3), 223–237.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224.

Hall, C. C., Ariss, L., & Todorov, A. (2007). The illusion of knowledge: When more information reduces accuracy and increases confidence. *Organizational Behavior and Human Decision Processes*, *103*(2), 277–290.

Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement,

15206793, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mar.21866 by Edinburgh University, Wiley Online Library on [12/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

15206793, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mar.21866 by Edinburgh University, Wiley Online Library on [12/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and business outcomes: A meta-analysis. *Journal of Applied Psychology*, *87*(2), 268–279.

He, Q., Glas, C. A. W., Kosinski, M., Stillwell, D. J., & Veldkamp, B. P. (2014). Predicting self-monitoring skills using textual posts on Facebook. *Computers in Human Behavior*, *33*, 69–78.

Hess, M., & Doe, P. (2013). The marketer's dilemma: Focusing on a target or a demographic?: The utility of data-integration techniques. *Journal of Advertising Research*, *53*(2), 231–236.

Homer, P. M. (1995). Ad size as an indicator of perceived advertising costs and effort: The effects on memory and perceptions. *Journal of Advertising*, *24*(4), 1–12.

Hulland, J., & Miller, J. (2018). Keep on turkin? *Journal of the Academy of Marketing Science*, *46*, 789–794.

Ioannou, A., Tussyadiah, I., & Marshan, A. (2021). Dispositional mindfulness as an antecedent of privacy concerns: A protection motivation theory perspective. *Psychology & Marketing*, *38*(10), 1766–1778.

Jaworski, B. J., & Kohli, A. K. (1993). Market orientation: Antecedents and consequences. *Journal of Marketing*, *57*, 53–70.

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An analysis of data quality: Professional panels, student subject pools, and Amazon's Mechanical Turk. *Journal of Advertising*, *46*(1), 141–155.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, *54*(3), 241–251.

Kohli, A. K., & Jaworski, B. J. (1990). Market orientation: The construct, research propositions, and managerial implications. *Journal of Marketing*, *54*(2), 1–18.

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality*, *64*(2), 311–337.

Kruglanski, A. W. (1989). The psychology of being "right": The problem of accuracy in social perception and cognition. *Psychological Bulletin*, *106*(3), 395–409.

Kumar, V., Rajan, B., Salunkhe, U., & Joag, S. G. (2022). Relating the dark side of new-age technologies and customer technostress. *Psychology & Marketing*, *39*(12), 2240–2259.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, *52*(2), 21–31.

Lefkeli, D., Tulan, D., & Gürhan-Canli, Z. (2022). Being observed in the digital era: Conceptualization and scale development of the perception of being observed. *Psychology & Marketing*, *39*(10), 1992–2008.

Li, F., Larimo, J., & Leonidou, L. C. (2021). Social media marketing strategy: Definition, conceptualization, taxonomy, validation, and future agenda. *Journal of the Academy of Marketing Science*, *49*(1), 51–70.

Lim, W. M., Kumar, S., Verma, S., & Chaturvedi, R. (2022). Alexa, what do we know about conversational commerce? Insights from a systematic literature review. *Psychology & Marketing*, *39*(6), 1129–1155.

Lyngdoh, T., El-Manstrly, D., & Jeesha, K. (2023). Social isolation and social anxiety as drivers of generation Z's willingness to share personal information on social media. *Psychology & Marketing*, *40*(1), 5–26.

Lyon, D. (2007). *Surveillance studies: An overview*. Polity Press.

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, *15*(4), 336–355.

Marder, B. (2018). Trumped by context collapse: Examination of 'Liking' political candidates in the presence of audience diversity. *Computers in Human Behavior*, *79*, 169–180.

Martin, K. D., Borah, A., & Palmatier, R. W. (2017). Data privacy: Effects on customer and firm performance. *Journal of Marketing*, *81*(1), 36–58.

Maseeh, H. I., Jebarajakirthy, C., Pentecost, R., Arli, D., Weaven, S., & Ashaduzzaman, M. (2021). Privacy concerns in e-commerce: A multilevel meta-analysis. *Psychology & Marketing*, *38*(10), 1779–1798.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, *90*(10), 60–66.

McKechnie, S., Nath, P., & Xun, J. (2018). New insights into emotion valence and loyalty intentions in relational exchanges. *Psychology & Marketing*, *35*(2), 160–169.

Micu, A., Micu, A. E., Geru, M., & Lixandroiu, R. C. (2017). Analyzing user sentiment in social media: Implications for online marketing strategy. *Psychology & Marketing*, *34*(12), 1094–1100.

Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing*, *58*(3), 20–38.

Moriuchi, E. (2021). An empirical study of consumers' intention to use biometric facial recognition as a payment method. *Psychology & Marketing*, *38*(10), 1741–1765.

Moriuchi, E. (2023). "Alexa, lock my front door": An empirical study on factors affecting consumer's satisfaction with VCA-controlled security devices. *Psychology & Marketing*, *40*(1), 169–189.

Norberg, P. A., & Horne, D. R. (2007). Privacy attitudes and privacy-related behavior. *Psychology & Marketing*, *24*(10), 829–847.

Okazaki, S., Eisend, M., Plangger, K., de Ruyter, K., & Grewal, D. (2020). Understanding the strategic consequences of customer privacy concerns: A meta-analytic review. *Journal of Retailing*, *96*(4), 458–473.

Okazaki, S., Li, H., & Hirose, M. (2009). Consumer privacy concerns and preference for degree of regulatory control. *Journal of Advertising*, *38*(4), 63–77.

Park, M. S., Shin, J. K., & Ju, Y. (2015). A taxonomy of social networking site users: Social surveillance and self-surveillance perspective. *Psychology & Marketing*, *32*(6), 601–610.

Pitt, C. S., Bal, A. S., & Plangger, K. (2020). New approaches to psychographic consumer segmentation: Exploring fine art collectors using artificial intelligence, automated text analysis and correspondence analysis. *European Journal of Marketing*, *2*, 305–326.

Pitt, C. S., Plangger, K. A., Botha, E., Kietzmann, J., & Pitt, L. (2019). How employees engage with B2B brands on social media: Word choice and verbal tone. *Industrial Marketing Management*, *81*, 130–137.

Plangger, K., Grewal, D., de Ruyter, K., & Tucker, C. (2022). The future of digital technologies in marketing: A conceptual framework and an overview. *Journal of the Academy of Marketing Science*, *50*, 1125–1134.

Plangger, K., & Montecchi, M. (2020). Thinking beyond privacy calculus: Investigating reactions to customer surveillance. *Journal of Interactive Marketing*, *50*, 32–44.

Plangger, K., & Watson, R. T. (2015). Balancing customer privacy, secrets, and surveillance: Insights and management. *Business Horizons*, *58*(6), 625–633.

Potoglou, D., Dunkerley, F., Patil, S., & Robinson, N. (2017). Public preferences for internet surveillance, data retention and privacy enhancing services: Evidence from a pan-European study. *Computers in Human Behavior*, *75*, 811–825.

Qi, J., Zhang, Z., Jeon, S., & Zhou, Y. (2016). Mining customer requirements from online reviews: A product improvement perspective. *Information & Management*, *53*(8), 951–963.

Rehman, Z. U., Baharun, R., & Salleh, N. Z. M. (2020). Antecedents, consequences, and reducers of perceived risk in social media: A systematic literature review and directions for further research. *Psychology & Marketing*, *37*(1), 74–86.

Rowe, F., Ngwenyama, O., & Richet, J. L. (2020). Contact-tracing apps and alienation in the age of COVID-19. *European Journal of Information Systems*, *29*(5), 545–562.

Salanova, M., Agut, S., & Peiró, J. M. (2005). Linking organizational resources and work engagement to employee performance and customer loyalty: The mediation of service climate. *Journal of Applied Psychology*, *90*(6), 1217–1227.

Scarpi, D., Pizzi, G., & Matta, S. (2022). Digital technologies and privacy: State of the art and research directions. *Psychology & Marketing*, *39*(9), 1687–1697.

Schweidel, D. A., Bart, Y., Inman, J. J., Stephen, A. T., Libai, B., Andrews, M., Rosario, A. B., Chae, I., Chen, Z., Kupor, D., Longoni, C., & Thomaz, F. (2022). How consumer digital signals are reshaping the customer journey. *Journal of the Academy of Marketing Science*, *50*, 1257–1276.

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly*, *35*(4), 989–1016.

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information privacy: Measuring individuals' concerns about organizational practices. *MIS Quarterly*, *20*(2) 167–196.

Sweeney, J. C., & Soutar, G. N. (2001). Consumer perceived value: The development of a multiple item scale. *Journal of Retailing*, *77*(2), 203–220.

Thomsen, E. (2002). *OLAP solutions: Building multidimensional information systems*. John Wiley & Sons.

Trivedi, R. H., & Teichert, T. (2021). Consumer reactions to nudity in print advertising: Comparing same-gender and opposite-gender effects. *Psychology & Marketing*, *38*(5), 819–833.

Turow, J. (2008). *Niche envy: Marketing discrimination in the digital age*. MIT Press.

Uleman, J. S. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology*. Guildford Press.

Visentin, M., Tuan, A., & Di Domenico, G. (2021). Words matter: How privacy concerns and conspiracy theories spread on twitter. *Psychology & Marketing*, *38*(10), 1828–1846.

Watson, R. T. (2013). *Data management: Databases and organizations*. eGreen Press.

Watson, R. T., Plangger, K., Pitt, L., & Tiwana, A. (2023). A theory of information compression: When judgments are costly. *Information Systems Research*. https://doi.org/10.1287/isre.2022.1163

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, *80*(6), 97–121.

How to cite this article: Plangger, K., Marder, B., Montecchi, M., Watson, R., & Pitt, L. (2023). Does (customer data) size matter? Generating valuable customer insights with less customer relationship risk. *Psychology & Marketing*, 1–13. https://doi.org/10.1002/mar.21866

APPENDIX A: Measurement scales used in experiments

Personality Scale (used in Study 1)	
Source:	Adapted from Gilbert and Warren (1995)
Type:	Likert 7 Point (Strongly disagree: Strongly agree)
Prompt:	From the personal data you have seen above, please answer these questions about that person's [brand name] habits
Items:	Economizer
	This person shops a lot for specials
	This person thinks that they can save a lot of money by shopping around for specials
	This person usually watches the advertisements for announcements of sales
	This person find themselves checking the prices in the grocery store even for small items
	Credit user
	This person likes to pay cash for everything they buy*
	This person buys many things with a credit or charge card
	This person think that to buy anything, other than a house or a car, on credit is unwise*
	This person thinks that it is good to have charge accounts
	Self confident
	This person is more independent than most people
	This person thinks they have more self-confidence than most people
	This person thinks they have a lot of personal ability
	Home oriented
	This person would rather spend a quiet evening at home than go to a party
	This person is a homebody
	This person likes parties where there is lots of music and talk*
	Fashionable
	When this person must choose between the two, they usually dress for fashion, not for comfort
	This person usually has one or more outfits of the very latest styles
	An important part of this person's life and activities is dressing smartly
	Note: Reserve coded items are denoted by an asterisk (*)
Brand attitude (used in Studies 2 and 3)	
Source:	Adapted from Homer (1995)
Type:	Bipolar 9 point
Prompt:	Please express your attitude toward ___________
Items:	Negative: Positive
	Unpleasant: Pleasant

15206793, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mar.21866 by Edinburgh University, Wiley Online Library on [12/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

	Disagreeable: Agreeable
	Worthless: Valuable
	Bad: Good
	Foolish: Wise
	Unfavorable: Favorable
	Dislike a lot: Like a lot
	Useless: Useful
Purchase likelihood (used in Studies 2 and 3)	
Source:	Item developed by Author team
Type:	Likert 7 point (Strongly disagree: Strongly agree)
Prompt:	From the personal data you have seen above, please answer these questions about that person's [brand name] habits
Items:	This person would be willing to buy [brand name] within the next year

15206793, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mar.21866 by Edinburgh University, Wiley Online Library on [12/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License