AMERICAN MARKETING ASSOCIATION

Article

Journal of Marketing
2021, Vol. 85(1) 163-183

DOI: 10.1177/0022242920969198
journals.sagepub.com/home/jmx
SAGE

Capturing Marketing Information to Fuel Growth

Rex Yuxing Du, Oded Netzer, David A. Schweidel, and Debanjan Mitra

Editor's Note: This article is part of the JM-MSI Special Issue on "From Marketing Priorities to Research Agendas," edited by John A. Deighton, Carl F. Mela, and Christine Moorman. Written by teams led by members of the inaugural class of MSI Scholars, these articles review the literature on an important marketing topic reflected in the MSI Priorities and offer an expansive research agenda for the marketing discipline. A list of articles appearing in the Special Issue can be found at http://www.ama.org/JM-MSI-2020.

MSI MARKETING SCIENCE INSTITUTE

Abstract

Marketing is the functional area primarily responsible for driving the organic growth of a firm. In the age of digital marketing and big data, marketers are inundated with increasingly rich data from an ever-expanding array of sources. Such data may help marketers generate insights about customers and competitors. One fundamental question remains: How can marketers wrestle massive flows of existing and nascent data resources into coherent, effective growth strategies? Against such a backdrop, the Marketing Science Institute has made "capturing information to fuel growth" a top research priority. The authors begin by discussing the streetlight effect—an overreliance on readily available data due to ease of measurement and application—as contributing to the disconnect between marketing data growth and firm growth. They then use the customer equity framework to structure the discussion of six areas where they see substantial undertapped opportunities: incorporating social network and biometric data in customer acquisition, trend and competitive interaction data in customer development, and unstructured and causal data in customer retention. The authors highlight challenges that obstruct firms from realizing such data-driven growth opportunities and how future research may help overcome those challenges.

Keywords

analytics, biometrics, competitive intelligence, field experiments, growth, social network, text analysis, trendspotting

Marketing is the functional area primarily responsible for driving a firm's organic growth. With increasingly rich data from an ever-expanding array of sources, marketers can now capture abundant information to derive more actionable insights on customers and competitors. These insights can fuel firm growth, yet there is also the potential for clutter, confusion, and misuse. Indeed, the CMO Survey (2020) reports a low level of contribution of marketing analytics to firm performance and no improvement in that contribution over the last eight years. This raises the question of how marketers can wrestle massive data flows into information relevant for effective growth strategies, turning data into a driver of long-term growth. The Marketing Science Institute (MSI) has made tackling this challenge a top priority.

To address the disconnect between marketing data growth and firm growth, we recognize that the use of new data is not, in and of itself, a growth strategy. Rather, a firm's marketing data and application mix must align with its growth strategy and, in doing so, provide the firm with a powerful strategic capability (Davenport and Harris 2017). However, evidence suggests that firms' data and application mix may not always align with resource allocations across growth strategies. The "streetlight effect" is one threat to this alignment, which stems from managers' overreliance on readily available data due to ease of

Rex Yuxing Du is Bauer Professor of Marketing, University of Houston, USA (email: rexdu@bauer.uh.edu). Oded Netzer is Arthur J. Samberg Professor of Business, Columbia University, USA (email: onetzer@gsb.columbia.edu). David A. Schweidel is Professor of Marketing, Emory University, USA (email: dschweidel@emory.edu). Debanjan Mitra is Voya Financial Chair and Professor of Marketing, University of Connecticut, USA (email: debanjan.mitra@uconn.edu).

measurement and application, irrespective of their growth objective. The streetlight effect manifests in managers leveraging "big data" for small problems while making do with small data for big problems or, even worse, neglecting data altogether. For example, the streetlight effect caused by the abundance of advertising data leads managers to focus more on managing advertising than on distribution or product line, despite advertising elasticities being much lower than the elasticities of the latter two (Hanssens and Pauwels 2016). Likewise, the streetlight effect of abundant near-market knowledge leads firms to choose proximity over favorable growth opportunities in distant international markets (Mitra and Golder 2002).

To guard against the streetlight effect, marketers must explicate how a firm's growth strategy can be supported by its portfolio of marketing data and applications. The objective is to prioritize data and applications to maximize the value of the decisions they support.

Relying on the premise that the "valid definition of business purpose (is) to create a customer" (Drucker 1954, p. 37), we suggest that the pertinent value must relate to customers. Blattberg and Deighton (1996) describe growth in terms of maximizing the value of the customer base, which arises from three drivers of customer equity: (1) *acquiring* new customers, (2) *developing* existing customer relationships, and (3) *retaining* customers (Gupta, Lehmann, and Stuart 2004). Accordingly, we frame the quest for marketing data to fuel firm growth along these three customer equity dimensions.

The remainder of this article proceeds as follows. We first look at historical examples of the streetlight effect as it pertains to marketing data. We then structure our discussion of opportunities for marketing data–driven growth around the three customer equity components (customer acquisition, customer development and customer retention), highlighting how undertapped data sources offer growth opportunities. For each component, we discuss the challenges that can obstruct such opportunities and identify directions for future research. The opportunities that we discuss—the use of biometric and social network data in customer acquisition, trend and competitive interaction data in customer development, and unstructured and causal data in customer retention—are by no means comprehensive, but we hope they can spark future work that leverages data with growth objectives in mind.

Marketing Data and the Streetlight Effect

Historical Marketing Examples of the Streetlight Effect

Data innovations are often thrust on marketers, who then scramble to leverage these data for different purposes using a variety of applications that unleash a firehose of information but may not always contribute to the firm's growth in the long run. This is one important contributor of the streetlight effect. In this subsection, we review several prominent examples of past marketing data innovations that demonstrate the streetlight effect and the resultant blind spots.

Retail scanner/scanner panel data. The advent of retail scanner and scanner panel data was a major breakthrough in the 1980s and 1990s (e.g., Guadagni and Little 1983). However, as Lodish and Mela (2007) caution, the increasing prevalence of real-time transaction data has made firms more myopic in their search for growth, overinvesting in promotions because the short-term effects are readily quantifiable and underinvesting in brand equity, new products, and distribution because their long-term effects are difficult to assess.

CRM data. With the spread of loyalty programs and advances in database technology, the advent of all-encompassing customer relationship management (CRM) data has enabled marketers to track the entire history of interactions with their own customers. However, most CRM data contains little information on the interactions that customers have with competitors or how customers' needs and wants evolve over time (Du, Kamakura, and Mela 2007). Such an imbalance between data on internal versus external relationships can lead to growth strategies that are overly inward- and backward-looking.

Clickstream data. Clickstream data that track customers across digital channels have afforded marketers a near 360-degree view of a customer's journey online. This has led to tools that support growth by improving marketing return on investment (ROI) through better target selection and media planning, including multitouch attribution models. However, touchpoints are not equally trackable across channels. Discrepancies in measurability between digital and mass media (e.g., TV, radio, print, outdoor) may have contributed to shifts of advertising budgets toward digital due to (1) larger measurement errors in mass media and (2) the difficulty in quantifying the generative influences of mass media and cross-media synergies.

Online testing/field experiment data. Recognizing the challenges inherent in drawing causal inferences from observational data, marketers increasingly turn to field experiments. Online testing has been the dominant driver behind the rise of field experimentation. Yet experimentation remains relatively rare when it comes to offline behaviors whose short-term responses are difficult to measure (e.g., Eastlack and Rao 1989) or marketing levers that are expected to have long-term effects (e.g., proactive retention efforts). Field experiments may also not capture the full range of marketing decisions, whether it be promotion amounts or ad copies, and their results may not be reliable over time. A relatively small set of domains where field experiments are prevalent, combined with increasing faith in them, could lead marketers to focus disproportionately on activities that are easy to manipulate and are more effective at driving measurable short-term responses, potentially at the expense of long-term growth.

User-generated content (UGC). Social media data and other forms of UGC have revolutionized the way marketers listen to their customers, dwarfing the data available through traditional tools such as surveys and focus groups. Popular UGC platforms use such data to target advertising. However, the size, timeliness, and richness of UGC does not guarantee that these data are representative. Relying on social listening could bias perceptions of the marketplace due to differences in users across platforms (Schweidel and Moe 2014; Schoenmueller, Netzer, and Stahl 2020).

Big data and machine learning methods. Many firms have jumped on the big data–machine learning bandwagon as open-source algorithmic methods become readily implementable, without fully grasping the relative advantages of traditional methods taken by marketing science or recognizing the potential for "algorithmic biases" and other unintended consequences (e.g., Lambrecht and Tucker 2019). In pursuing predictive ability, big data applications risk sacrificing the interpretability of the results. Consequently, marketers could inadvertently create social ills in their pursuit of growth that can harm society and, eventually, the firm.

Marketing Data Blind Spots

A common thread across these examples is that while new data sources can yield insights that lead to growth, they come with trade-offs. As marketers focus on particular data sources, our field of vision narrows, resulting in potential blind spots that can weaken a firm's growth trajectory, including the following.

First, marketing data may result in *prioritizing short-term growth ahead of long-term growth*. While some marketing efforts yield short-term responses, others must be assessed over a longer time frame. Because data for measuring short-term effects (e.g., click-through rates) are easier to obtain, data-driven growth may favor investments in lower-funnel actions (e.g., price promotions) at the expense of longer-term, upper-funnel actions (e.g., brand building).

Second, marketers may overly *rely on historical, internal data at the expense of forward-looking, external growth opportunities*. Firms have rich data about interactions with their customers. Growth strategies built on such data may favor customers with limited expansion potential while those whose relationships could be substantially enhanced are neglected.

Third, marketing data may create *a preference for more easily measured digital touchpoints at the expense of offline channels*. Despite the continued growth of e-commerce—especially during the COVID-19 pandemic—it accounts for only 14.5% of retail sales (eMarketer 2020). Given the prevalence of offline activities, efforts to leverage marketing data must balance online and offline sources.

Finally, marketers may *rely on available data in lieu of representative or predictive data*. This is exemplified by UGC, where research has cautioned about a vocal minority that may not represent the "silent majority" (e.g., Moe and Schweidel 2012). Another example of this blind spot arises from assuming, often incorrectly, that each data point is equally informative. Shugan and Mitra (2009, 2014) identify features in commonly encountered contexts where specific data points are more informative than others. These features result in the superiority of measures harnessed from selected data points (e.g., maximum, top rank displacement) as predictors of growth versus those that seek to capture information from all data points (e.g., average, variance).

In the next three sections, we outline specific undertapped data and research opportunities for each of the three drivers of customer equity. Tackling customer acquisition, development, and retention in turn, we discuss how the streetlight effect may have led to these opportunities being overlooked and how marketers may leverage them to fuel growth.

Customer Acquisition

Customer acquisition is critical to firm growth, as its success is necessary before a firm can focus on developing and lengthening the relationship. One way to achieve growth via acquisition is to use marketing data to increase the efficiency of acquisition efforts, such as by identifying prospects who are more likely to convert. While internal data on current customers may be rich, such data may contain limited information about prospects. Though customer profile data are readily available, there are questions as to their accuracy (Neumann, Tucker, and Whitfield 2019) and ability to capture prospects' state of mind at a given moment. Focusing on such available and limited data may lead marketers to overlook richer data sources. Next, we discuss two such sources that can provide timelier and often external insights into prospects that can improve acquisition efforts: biometric and social network data.

Incorporating Biometric Data into Customer Acquisition

In recent years, we have seen an explosion in biometric data being collected. Biometric data are physiological measures and calculations collected from individuals. Genetic testing kits from providers such as 23andMe offer insights into people's health, including screening for the BRCA1/BRCA2 mutations that indicate an increased risk of breast and ovarian cancer (U.S. Food and Drug Administration 2018). Beyond providing health insights, medical research is one use of genetic data (Nocera 2020). For business purposes, biometric data is being used to evaluate marketing creatives (e.g., Venkataraman et al. 2015), enabling marketing research firms to collect data on how individuals respond to advertising and identify creatives that are most likely to resonate with the target audience.

Marketers have examined an array of biometric data sources, including eye tracking (e.g., Pieters and Wedel 2004), electroencephalogram (e.g., Pozharliev et al. 2015), functional magnetic resonance imaging (e.g., Yoon, Gonzalez, and Bettman 2009), and emotion detection (e.g., Liu et al. 2018). A compelling aspect of biometric data is its real-time nature. Smartwatches and activity trackers monitor heart rate and blood pressure at a given moment. Such wearable devices also offer a means by which individuals can be motivated. For example, by incentivizing drivers to be monitored, insurance providers may find that delivering biometric information through wearables encourages the adoption of healthy habits.

Digital music providers and advertisers can coordinate the audio with a user's level of activity at the time.

While physiological responses to stimuli can offer insights into consumers in a controlled environment, there have been limited efforts to apply these insights for marketing in the field. The availability of consumer-facing products that collect biometric data and tools that can collect such data from large groups can enable marketers to incorporate such information to engage consumers in real time. Biometric data also offer opportunities for healthcare and technology providers to develop new services and acquire new customers (Cheng 2019). Collecting biometric responses to marketing content can enhance content personalization and product recommendation. Biometric data can also aid in contextual targeting to inform when messages should be delivered to support prospect acquisition.

Challenges in using biometric data. Biometric-based marketing research is being delivered by a range of providers, including Nielsen, Ipsos, and Kantar. In studies conducted by such providers, participants are often exposed to stimuli and their physiological responses are monitored. Though such studies are powerful additions to the marketing researcher's toolkit, the technology lacks portability. While functional magnetic resonance imaging and electroencephalogram studies can monitor how neurological activity is affected by external stimuli, such studies are often expensive and must occur in a laboratory environment (Venkataraman et al. 2015).

Collecting biometric data outside the lab. One of the key challenges in the use of biometric data to support customer acquisition is the ability to collect such data both at scale and in the field. Connected devices such as smartwatches and activity trackers may facilitate field data collection, enabling the capture of heart rate, blood pressure, and the rate of oxygen consumption. Some techniques, such as eye tracking and emotion detection, can already be deployed in the field. Once the data collection challenge has been addressed, marketers can investigate the relationship between biometric measures and acquisition to better understand the stimuli and context in which different biometric measures arise. Doing so can then support prospect identification and content development to increase the efficiency of acquisition efforts.

Navigating consumer privacy concerns. Marketers must be cautious when harvesting personal information from individuals. Biometric data may be used for unintended purposes, such as facial recognition using data scraped from social media posts or retailers using video surveillance data to identify engaged prospects without obtaining their explicit consent (Puntoni et al. 2021). Not only are consumers likely to vary in their privacy preferences, they may also react differently to data protection measures such as anonymization or aggregation.

Opportunities and future research directions in using biometric data. The opportunities for biometric data to support growth require that metrics be collectable in the field. Wearables provide a convenient means by which such data can be collected. Smartwatches can collect measures such as heart rate and oxygen consumption levels. The use of such data by marketers depends on its being made available by manufacturers. In discussing growth opportunities using biometric data, we focus on eye tracking (e.g., Pieters and Wedel 2004) and emotion detection (e.g., Liu et al. 2018), which can be implemented at scale. Methods are being developed and improved to detect emotions from individuals in a crowd (e.g., Favaretto et al. 2019). However, we have yet to see these tools being applied to large crowds. Challenges remain in using eye tracking in the wild due to each individual's "idiosyncratic field of view at each point in time during the recording" (Bulling and Wedel 2019, p. 38).

Optimizing marketing messages. Retailers, hospitality businesses, and entertainment venues could make use of biometric data to target and optimize marketing messages. Exterior signage equipped with video cameras would enable marketers to identify the characteristics of messages that are resonating with passersby based on the emotions they express and if the content is grabbing their attention. Identifying the content that both attracts attention and elicits a positive emotional response may better engage prospects and increase conversion. Another direction for research is the optimization of marketing content delivery timing. Using eye tracking and facial detection systems, future research could examine how to optimize the schedule with which content is shown to large groups to increase acquisition. While this would ideally involve linking an individual who is exposed to multiple marketing messages to their subsequent behavior, which could be achieved through facial recognition, researchers should work toward privacy-friendly methods to connect media exposures and subsequent activities.

Identifying biometric responses throughout the customer journey. Because one of the advantages of biometric data is its real-time availability, research could use such data to examine how consumers physiologically react to marketing throughout the customer journey. At different stages, certain biometric measures may be more informative of the likelihood of progressing to the next stage. In particular, how do consumers react to marketing in the steps leading up to the acquisition decision? In the prepurchase stage that precedes acquisition, consumers engage in need recognition, consideration, and search (Lemon and Verhoef 2016). Eye tracking may inform consumers' consideration sets and search processes based on the products viewed and the features that attract attention. The emotions expressed when different products are viewed can inform subsequent product recommendations. Research should also consider the intensity of emotions that consumers express. Real-time data collection can also enable analysts to examine how variations in biometric measures (e.g., difference, velocity, acceleration) relate to acquisition behavior, which can enable marketers to target consumers at the times when they would be most receptive to acquisition efforts. Emotional trajectories likely vary across categories, perhaps with consumers exhibiting more emotional variation in response to experiential and hedonic products while they may not express as much emotional variation in their search for and consideration of utilitarian products. This may suggest that eye tracking measures and emotional measures have varying degrees of informativeness on the likelihood of acquisition in different categories.

Using biometrics to measure the effectiveness of media and embedded marketing. Biometric data can also create new opportunities to understand the drivers of acquisition by capturing physiological measures during media consumption activities (or interactions with salespeople in a business-to-business [B2B] context). An individual's biometric responses can be gathered while they are exposed to videos or music on mobile devices, whether it is through a smartwatch or earbuds that collect measures such as heart rate. Such biometric measures will fluctuate based on the media being consumed, including the focal content and embedded marketing. Tonietto and Barasch (2020) report that media consumers who generate content related to the consumption experience report increased immersion and engagement. The extent of consumers' immersion may be measurable in real time using biometric data. Moreover, researchers can use biometric data to identify when consumers will be most open to marketing efforts and have an increased tendency for conversion, such as high arousal and positive emotion. These measures may be informative of purchase intent, offering marketers a time window during which to deliver incentives.

In addition to the effects of embedded marketing messages, researchers using biometric data may consider the sequence of emotional reactions that are produced along the path to purchase, and whether these reactions are affected by marketing messages or the content in which they are embedded. If consumers are immersed in the content (e.g., Hoffman and Novak 1996), outside factors such as marketing messages may disrupt the consumption experience and elicit negative reactions that adversely affect the brand. By collecting biometric response throughout media consumption, such misfires can potentially be avoided.

Incorporating the Social Network into Customer Acquisition

While methods for customer (Lewis 2006; Schweidel, Bradlow, and Fader 2011) and corporate (e.g., Gupta, Lehmann, and Stuart 2004; McCarthy, Fader, and Hardie 2017) valuation that account for customer acquisition have been developed, they are not without shortcomings. Such models often are constructed on the assumption that customers act independently. While this lowers data requirements and offers computational benefits, this assumption may also limit the accuracy of the valuation model.

Word-of-mouth (WOM) activity plays a critical role in customer acquisition. Trusov, Bucklin, and Pauwels (2009) report that WOM activity is often more impactful than traditional marketing actions in acquiring users. Villanueva, Yoo, and Hanssens (2008) find that customers acquired through WOM deliver more long-term value than those acquired through traditional marketing. Others have demonstrated similar findings regarding offline WOM activities (e.g., Bell and Song 2007). The source of WOM has also been identified as an important consideration (e.g., Iyengar, Van den Bulte, and Valente 2011).

One way in which firms can observe social ties is through customer referrals (e.g., Kumar, Petersen, and Leone 2010). However, referrals reveal only a portion of the social ties. Secondary data sources can provide a more complete picture. While social interactions are ubiquitous, social ties and the information flow across these ties are difficult to observe, which may have limited their use (in comparison to referrals) for acquisition. Social ties data are often collected by third-party providers, further contributing to the streetlight effect that leads marketers to overlook such data. Another contributing factor to marketers ignoring this potentially fruitful data source is that, as we discuss subsequently, it can be difficult to directly interpret, leading to another manifestation of the streetlight effect.

Despite the challenges associated with using social network data, it offers a significant opportunity to drive growth via acquisition. In prioritizing prospects for acquisition efforts, focusing on a prospect's value based solely on their own expenditures neglects the value they provide through their impact on other prospects (e.g., Ho et al. 2012; Kumar 2018). Incorporating such data into acquisition efforts can help align the firm's data efforts with its growth objectives by leveraging interactions among members of the customer base.

Challenges to leveraging social network data. To incorporate social network structure into customer analytics and inform subsequent acquisition decisions, there are some fundamental challenges that must be addressed. We discuss two specific challenges as they pertain to customer acquisition.

Incomplete network data arising from data collection methods. Platforms for UGC have proven to be a rich data source, but limited research has documented differences across platforms (Schoenmueller, Netzer, and Stahl 2020; Schweidel and Moe 2014). As users' motives for engaging with each other likely differ across platforms, inferring social ties from one venue will not capture all meaningful connections between users. While LinkedIn connections may arise for professional reasons, those on Instagram may reflect personal or familial ties. The relevance of social ties must also be identified. A specialized forum for physicians or health care professionals could reveal interactions that are most important and the most influential nodes with regard to procedure adoption, but social ties on consumer-oriented venues such as Instagram or Pinterest may be useful for identifying prospects with similar socioeconomic status.

The streetlight effect has led to a preponderant focus on online (rather than offline) social ties and WOM. However, with most consumers still making purchases and interactions offline, offline social ties may be particularly relevant in certain industries. Though offline social networks can be inferred through sources such as mobile location data (e.g., Zubcsek, Katona, and Sarvary 2017), an overreliance on online social interaction data that are more readily available may bias ROI estimates of social network–based acquisition efforts (Chen, Chen, and Xiao 2013).

Failure to recognize the strength of social ties and their dynamics. Another challenge is the degree to which network ties and their strength are stable. The longevity of a connection and the frequency of interactions may inform the strength of the connection and the extent to which two individuals' decisions are correlated. Researchers often fall victim to the streetlight effect when they merely look at the existence of a social tie as opposed to the strength of the tie. It is not enough to observe the presence of ties, as data capturing the degree and direction of information flow along these ties is needed to infer the relevance of the relationship.

A related challenge to using social network data in customer acquisition is the shelf life of the data. While consumers add online connections, they rarely take steps to prune them. If individuals do not actively sever social ties, we observe a situation akin to latent attrition, with no activity occurring along a social tie for an extended period of time. Acquisition efforts require up-to-date information on the social ties and their strength that may influence prospects' choices, requiring that marketers distinguish between active and dormant social ties.

There also remain technical challenges associated with capturing and analyzing dynamic social network data. While dyad-level data can characterize the relationship between consumers, the data grow quadratically with the size of the network (Braun and Bonfrer 2011). Adding a temporal dimension to the interactions results in rapid growth of the data, which may require marketers to predetermine the prospects that warrant their attention in their network analysis.

Opportunities and future research directions in leveraging social network data. Having identified the challenges pertaining to missing data in the construction of the social network and the dynamic nature of social ties, we now discuss specific opportunities that we see to leverage social network structure to drive growth via better targeted acquisition efforts.

Reconceptualizing customer value. Ho et al. (2012) discuss the influencer value of a customer, recognizing that a customer's social influence may induce the acquisition of others. Gupta and Mela (2008) recognize the impact of direct network effects in customer value, as the presence of nonpaying customers may help attract paying customers. In the context of a two-sided platform, Yao and Mela (2008) estimate the forgone value associated with an individual. Future research can reevaluate the way in which we consider the value that a customer delivers to an organization. Recognizing the knowledge value that individuals provide (e.g., Kumar 2018), research can view interactions between prospects as opportunities for knowledge to be shared, which may either be retransmitted to other prospects or directly affect a prospect's decision. The volume and content of such interactions allow the experiences of a single prospect to affect others, thereby affecting perceptions and future expectations of customer acquisition.

In the case of acquisition, one may consider the forgone value of an individual prospect. Not only are the expenditures by the prospect lost, but so is the value from others that the focal prospect may have affected. The value of some individuals, depending on their position in the network, may arise primarily through their impact on others. Though social effects have been probed in the context of customer retention (e.g., Ascarza et al. 2017), to the best of our knowledge, there have been no efforts to consider it with regard to acquisition.

A related opportunity is to derive the forgone value associated with combinations of (possibly socially connected) customers. Deriving the value of customer combinations upends traditional prospect scoring. Among the research opportunities that this creates is the development of techniques to evaluate the impact of different resource allocations across prospects. We expect that individuals with strong social ties have the ability to influence others' acquisition decisions. We also expect a stronger impact on customer acquisition in contexts where network density is high, as messages may propagate further. Developing a joint model for information flow through a prospect network and acquisition decisions could help evaluate counterfactuals. For instance, should the firm reduce its efforts on tightly connected prospects? This may be feasible if the firm can leverage social interdependence to achieve the same rate of acquisition using fewer resources.

Given the number of combinations of prospects, it may not be feasible to derive the forgone value of a unique prospect. One alternative may be to characterize prospects using internal classifications aligned with operations (e.g., marketing territories) to estimate the forgone value associated with one (or more) prospects of a certain type. Identifying actionable proxies for the social network structure and incorporating them into resource allocation models is another area in which future research is warranted.

Supplementing third-party data. Another opportunity lies in comparing the value of various external social interaction data sources for identifying high potential prospects. Neumann, Tucker, and Whitfield (2019) report considerable variation in the veracity of third-party profiles. Marketers' reliance on such data sources to identify prospects could be improved if they can be merged with data on the social interactions of their current customers. Prospects' positions in the social network relative to current customers could inform which prospects warrant attention. By using social network data to screen prospects, marketers may increase conversion, boosting the ROI of their efforts and driving growth. One of the reasons that social network data can supplement third-party profiles is the ability to focus on activity within a particular time frame. As social interaction data include a timestamp, marketers may focus on recent interactions. If consumers typically conduct product research for a few weeks prior to making a purchase in a category, marketers may focus on the social connections with which an individual has interacted during this timeframe.

The benefits of using social network data must be weighed against privacy concerns (Cui et al. 2021). While consumers may accept the use of third-party profiles in gauging their purchase intent, they may react adversely to firms compiling their social ties. To avoid consumers' backlash, marketers must exercise caution in sourcing social interaction data, as well as in the extent to which marketing communications reveal the rationale by which prospects have been selected for targeting (e.g., Hersh and Schaffner 2013). Network typology as reflected in node-level summary statistics may be an alternative to the use of detailed social ties that can mitigate consumers' privacy concerns. Other metrics that may be informative of a prospect's likelihood of becoming a customer include the fraction of social ties who are current customers or the average expenditures of an individual's social connections. While the former captures the extent to which the social ties have a relationship with the firm, the latter may better reflect the monetary value of these social ties. Research can identify privacy-friendly measures that retain the predictive value of the social network with respect to acquisition tendencies.

Revisiting social selling. Another opportunity to use social network data for customer acquisition is social selling, a lead generation method whereby members of a sales force reach out to prospects on social media platforms. Social selling can be particularly effective in the context of B2B marketing (Minsky and Quesenberry 2016), enabling marketers to expand their audience. Accounting for the network structure among the individuals with which a firm engages is essential as it can affect the frequency with which prospects are exposed to content from the firm. Ignoring social connections may result in marketers saturating prospects with messages due to repeated exposure from social ties. In addition, in contrast to many consumer decisions, B2B purchases involve many individuals making a joint decision. The strength and density of the social connections among the individuals who comprise the decision-making team may inform the relationship between decision makers and the likely interactions among them.

Accounting for the network structure also has implications for sales force compensation. Salespeople should optimize their efforts across prospects, accounting for both expected sales to a prospect and their potential impact on others. The design of compensation plans should also minimize the potential for free riding off of the social effect generated by other salespeople's efforts. Developing compensation schemes that accommodate social interdependence in the presence of multiple salespeople is a promising direction for future research, which can better align the sales force and the company incentives, minimize undesirable behavior and serve as an opportunity to drive growth via acquisition in B2B settings.

Customer Development

Firms can grow by expanding the relationship with their existing customers. Data-driven approaches to customer development often rely on internal data on customer–firm interactions to identify targeting and personalization opportunities. Such data are readily available and have been widely used to build applications such as recommender systems for cross- and upselling. However, an overreliance on internal customer–firm interaction data for customer development can make a firm vulnerable to the streetlight effect, risking backward- and inward-looking myopia. Backward-looking myopia arises as customer needs and wants change due to shifting trends, which make past purchases less predictive of future preferences. Inward-looking myopia stems from customers interacting with multiple firms, with internal data providing only a partial view of a customer's category demand. In this section, we focus on two undertapped data sources—trend and competitive interaction—that can help mitigate these risks, leading to new growth opportunities through more forward- and outward-looking customer development efforts.

Leveraging Trend Data in Customer Development

Changing customer needs and wants can manifest in shifting preferences for existing product features (Du, Hu, and Damangir 2015) or emerging demands for new features that can reshape well-established product category boundaries and cost–benefit trade-offs in customers' purchase decisions (Ofek and Wathieu 2010). Spotting trends in the evolution of customer mindsets and behaviors before the competition and adjusting the marketing strategies of existing offerings or developing new offerings are key opportunities for growth through customer development. The data that inform trends are often external, mitigating the risk of inward-looking myopia. In addition, trend data (by definition) focus on longer-term growth as opposed to short-term gains. Despite the importance of trendspotting to inform growth opportunities, marketing scholars have provided little guidance on how to systematically gather and analyze data to generate such foresights (Du and Kamakura 2012). Several challenges remain.

Challenges with trendspotting

Separating emergent trends from fads. The biggest challenge in trendspotting is that, without the benefit of hindsight, it is difficult to distinguish short-lived fads from emergent trends that offer meaningful growth opportunities. Consequently, it is risky to implement a proactive growth strategy contingent on identifying emerging trends in customer needs.

Take gluten-free and aspartame-free as two contrasting examples. Gluten-free foods have grown from a niche category into a multibillion dollar business where avoiding gluten is now a lifestyle choice (Mintel 2018), despite the lack of scientific evidence for its benefits (Levinovitz 2015). Gluten-free products have become a source of sustained growth for many consumer packaged goods manufacturers (Packaged Facts 2016). By contrast, PepsiCo's recent decision to remove aspartame (an artificial sweetener) from its flagship Diet Pepsi turned out to be a debacle akin to the 1985 "New Coke" fiasco, having to be pulled and replaced with the original after three years of declining sales (Kelso 2018). The root cause of PepsiCo's blunder is that it mistook a fad (health concerns about aspartame) for a trend, despite the due diligence and market research accompanying such a high-stake decision (Ester and Mickle 2015).

Synthesizing across domains and sources. Another challenge with identifying relevant trends is that they often transcend industries, technologies, and products, representing major shifts in consumer mindsets and behaviors that arise from social, economic, environmental, political, and technological changes (Ofek and Wathieu 2010). Identifying such transcendental trends requires surveying and integrating a multitude of sources across domains (e.g., expert interviews, trade publications, industry reports), which can often diverge from or even contradict with one another. Such cross-domain, cross-source synthesis presents a major methodological hurdle.

Identifying relevant data sources and metrics. A practical trendspotting challenge is the collection of archival data to quantify the magnitude and momentum of trends and fads at any given moment in time, including the pretakeoff period. Because historical sales data could be difficult to obtain for a large number of trends and fads, an alternative data source could be Google Trends (trends.google.com/trends/), which can be used to gauge Google users' interest in virtually any topic (Du and Kamakura 2012; Du, Hu, and Damangir 2015). Keyword search volume by market can be obtained dating back to 2004, which has been shown to be correlated with sales (Du and Kamakura 2012).[1] Researchers can also gather historical data from social listening platforms (e.g., Brandwatch, Meltwater) to quantify shifts in consumer interests based on the volume and content of online posts, which have been shown to be informative of consumers' mindset and behavior (Schweidel and Moe 2014; Zhong and Schweidel 2020).

Beyond natively digital sources such as search and social media, other sources of text can prove useful for trendspotting. Researchers can take advantage of digitized texts that have appeared in books, periodicals, patents, and other publications (e.g., Watts 2018). For example, researchers can now easily access a corpus of digitized texts containing 4% of all books ever printed (https://books.google.com/ngrams). The data set spans over 200 years (1800 through 2019) and comprises yearly data on occurrence frequency of two billion one-through five-n-grams.[2] Michel et al. (2011) illustrate how such data enable quantitative investigations of cultural trends, generating insights about fields as diverse as technology adoption, lexicography, collective memory, the pursuit of fame, censorship, and epidemiology.

Opportunities and future research directions in leveraging trend data. Beyond compiling the necessary data sources for trendspotting, new econometric or machine learning methods for conducting large-scale trend analysis are needed to better separate emergent trends from fads as early as possible and to predict long-term trend trajectories. To that end, methodological advances can be made in several fronts.

Synthesizing multiple data sources. For each historical trend or fad, there can be many relevant keywords and topics manifesting across many sources (e.g., online searches, posts in different social media platforms, mentions in different news media). One challenge to integrating different data sources to derive a meaningful index at any period in time lies in their varying periodicities. While some measures can be collected on a daily or weekly basis, others may only be available at a monthly (e.g., trade publications), quarterly (e.g., financial reports), or yearly (e.g., letters to shareholders) level. Research is needed to synthesize trend data with disparate reporting frequencies to isolate the common underlying signal (e.g., Ascarza and Hardie 2013) while accounting for other intrinsic differences that may exist across sources (e.g., search volumes may be a stock variable of consumer interest while social media mentions may be a flow variable, UGC may capture shifts in demands while company press releases may capture shifts in the way that managers view the marketplace).

Another methodological opportunity for researchers is the development of methods to integrate both structured and unstructured data in deriving trend indexes. Structured data such as the volume of keyword searches or topic mentions in textual social media posts can be combined with unstructured data such as nontextual social media posts to allow researchers to go beyond textual data and tap into visual, audio, and video data to spot trends that can drive growth through customer development (Berger et al. 2020).

Conducting early long-range forecasting. Projecting the trajectory of a trend or fad is a fundamental challenge in time series modeling, especially during the pretakeoff period when data are limited and accurate long-range forecasts are highly valuable (Chandrasekaran and Tellis 2007). One way to solve this "cold start" problem is to build a large training sample of historical trends and fads that cover a wide range of industries, pairing it with pattern recognition methods to identify the most similar historical counterparts to help make forecasts for the emergent trend or fad (Heist and Tarraf 2016). Unfortunately, no such a training sample exists, the construction of which will require painstaking efforts. To do so, one may follow the principles that have been successfully applied to discover empirical regularities in the diffusion of technological innovations and new products (e.g., Golder, Shacham, and Mitra 2009).

Another direction that researchers may consider exploring is the identification of generalizable factors that can predict the trajectory of a trend or fad. In their application to customer base analysis, Dew and Ansari (2018) demonstrate how dynamic behavior can be decomposed into underlying components including calendar effects and individual-specific factors that affect transaction behavior. Researchers could consider a similar decomposition to the aforementioned data sources for better distinguishing trends from fads. In addition to the time at

[1] For example, Google Trends indexes for searches by U.S. consumers including the keyword "gluten" versus "aspartame" (https://trends.google.com/trends/explore?date=all&geo=US&q=gluten, https://trends.google.com/trends/explore?date=all&geo=US&q=aspartame).

[2] An n-gram is a contiguous sequence of n words from a given sample of text (see http://www.culturomics.org).

which the data are generated, researchers could also consider geographic variation based on the origin of the data. Deep learning methods such as long short-term memory networks may offer a useful tool, provided that an adequate training set is developed.

Identifying trendsetting markets or segments. The timing with which a trend or fad unfolds may vary across markets or segments, some of which could be "harbingers" that send early warning signals about what is to come (Anderson et al. 2015). Identifying these trendsetting markets or segments could help spotting the rise or fall of a trend or fad. For example, in which product categories do coastal or urban markets tend to lead inland or suburban markets, or vice versa? One promising data source for addressing such an empirical question about shopping trends would be Google Shopping Insights (https://shopping.thinkwithgoogle.com/), which provides daily shopping search data for 55,000+ products, 45,000+ brands, and nearly 5,000 categories, across all 210 designated market areas in the United States.

A related direction for future research is to investigate the nature of contagion among markets or segments. For instance, trends may spread via both online and offline WOM, with the latter relying more on geographic proximity (Bell and Song 2007). A geotemporal diffusion model that distinguishes between online (global) and offline (local) contagion may help identify how quickly a trend spreads both across and within markets, providing guidance to marketers on where and when their resources may best be deployed to increase customer expenditures.

Finally, thinking about the social ties discussed in the acquisition section, understanding the social ties in which a trend has emerged may also help assess the likelihood of its success (Dover, Goldenberg, and Shapira 2012). One way in which researchers may explore this opportunity would be to examine whether an increased focus on marketing to trendsetters among existing customers will increase expenditures by other customers with whom they have strong social ties.

Incorporating Competitive Intelligence Data into Customer Development

Firms face a fundamental information asymmetry at the customer level. While they observe their interactions with customers, they know little to nothing about a customer's interactions with competitors, resulting in an overreliance on readily available internal data. Obtaining customer-level competitive intelligence can enrich a firm's knowledge of both what a customer purchases from competitors and the offers the customer receives from competitors, which is essential to increasing the share of their category expenditures with the firm.

Challenges to customer-level competitive intelligence data

Assessing a customer's growth potential. Using internal data alone, a firm may misjudge a customer's growth potential because it cannot distinguish a customer with a small wallet of which it gets a large share from one with a large wallet of which it gets a small share. The latter may have more growth potential if the customer were correctly identified and targeted. Researchers have proposed methods for estimating size and share of customer wallet that comprise two main approaches: (1) augmenting internal transaction and customer characteristics data with external transaction data for a sample of customers (e.g., Du, Kamakura, and Mela 2007; Keiningham et al. 2011) using customer surveys or purchase panels run by syndicated data providers or third-party data aggregators, or (2) estimating size and share of customer wallet using only internal data (e.g., Chen and Steckel 2012). While the first approach may incur significant data acquisition costs, the latter approach requires imposing strong assumptions about the data generating processes and can only be validated indirectly due to the lack of external data.

Evaluating relationship expansion efforts. Firms also struggle to assess the effectiveness of relationship expansion efforts because they cannot observe the offers or counteroffers that customers receive from competitors. In the absence of such customer-level competitive intelligence, it is difficult to tailor offers that account for the competitive context for each customer. Few statistical and econometric methods are available for addressing the challenge of missing data on customer-level competitive activities (for an exception, see Moon, Kamakura, and Ledolter [2007]).

Conceptually, the pitfalls of ignoring competitive intelligence in managing customer relationships has long been recognized (e.g., Boulding et al. 2005). However, overcoming inward-looking myopia remains difficult for one primary hurdle: data on customers' activities with competitors have been difficult to obtain. The schism in availability between internal and external data, we contend, may have grown over time as firms increasingly track interactions with their own customers, increasing the detail and volume of data available to them, while customer-level competitive intelligence remains elusive due to increasingly personalized marketing efforts, media and channel fragmentation, and privacy concerns.

Opportunities and future research directions in customer-level competitive intelligence data.

The challenges described previously present growth opportunities for firms that can tackle the internal versus external information asymmetry to yield a competitive advantage in customer development, and promising directions for researchers to develop methodologies for the collection and analysis of customer-level competitive intelligence data.

Imputing size and share of wallet with multiple external data sources. While surveys are a common source for size and share-of-wallet data, self-reports can be time-consuming, costly, and error-prone. Keiningham, Buoye and Ball (2015) find that the rank ordering in perceptions and attitudes among brands used by a consumer tend to be highly predictive of a brand's share of wallet. Researchers may identify similar survey measures (e.g., consideration set, purchase intent, willingness to recommend) that can strike a balance between burden on the respondent and predictive ability. Once such measures prove cost-effective and reliable, they could be integrated into brand trackers. Beyond survey data, for a sample of customers firms may acquire size and share of wallet data from vendors that run purchase panels (e.g., comScore Networks, GfK, IRI, Kantar, Nielsen, NPD) or third-party aggregators that track customer transactions across competing vendors in a category (e.g., Acxiom, credit bureaus, IMS, IXI, Rakuten Intelligence).

Researchers may also try to integrate newer data sources that can be used to infer size and share of wallet. For example, just as social media mentions have been used to map brands' competitive positions (e.g., Netzer et al. 2012), researchers may explore the use of such data to infer brands' share of wallet at the customer level. Paired with customer transaction data, research may assess if the frequency and concentration of competitor mentions is informative of customers' size and share of wallet. Researchers may also explore the potential for clickstream data (via cookies installed by the firm's own website or through third-party aggregators) and mobile location data (e.g., Mogean, PlaceIQ). Specifically, to what extent do share of website and offline store visits relate to share of wallet? Collaborations with credit card panels such as Second Measure could provide a means of assessing such a relationship empirically. Wearables may also serve as tools to assess a customer's overall usage in some categories, or even the use of a specific competitor's products. For example, distance traveled recorded by a fitness tracker could be used to gauge demand for running shoes. Should a relationship between these newer data sources and share/size of wallet be borne out, such data could provide means for estimating the extent to which customers engage with competitors, mitigating the asymmetry caused by differences in the availability of internal and external transaction data.

Projecting the evolution of both size and share of wallet. Customer valuation and cross- and upselling models rely almost exclusively on internal data (Netzer, Lattin, and Srinivasan 2008; Schweidel, Bradlow, and Fader 2011), confounding the processes governing the evolution of wallet size and wallet share. When the former remains stable, existing models would be well suited as changes in internal transactions are driven mainly by changes in share of wallet. However, in situations where consumers' needs and preferences evolve over time (e.g., Du and Kamakura 2006), ignoring the distinct dynamics between wallet size and share could lead to erroneous assessments of growth potential.

Developing models that disentangle the evolution of wallet size and share remains an opportunity for future research. Changes to the wallet size are driven mainly by customers' purchasing power and needs, whereas shifts in the share of wallet are indicative of relationship strength. Own and competitive marketing efforts can affect both the size and share of wallet. For example, product development that incorporates new features by a single brand may spill over to other brands in a category by increasing the perceived category benefit. Exogenous events such as a global pandemic or brand crisis may affect both category demand and wallet share, depending on the strength of an individual's relationship with brands in the category. This could result in a form of double jeopardy, with shocks reducing both total category demand and the share of wallet of weaker brands. Modeling the joint evolution of these processes enables forecasting future expenditures for a given firm and for the entire category. Such efforts are nontrivial, as they must account for strategic behavior of the firm and its competitors.

Comparing predictive performance across models that (1) use only internal data and ignore the distinct dynamics between wallet size and wallet share; (2) use only internal data but account for the distinct dynamics between wallet size and wallet share; and (3) augment internal data with external data and account for the distinct dynamics between wallet size and wallet share could reveal the incremental value of each component, which could vary depending on the empirical context. For example, in categories where customer wallet sizes tend to vary substantially but predictably over time, it might be most important to model the dynamics of wallet size and wallet share separately, with or without external data. Purchase panel data that covers an extended time window (say, between five and ten years) and breaks out transactions by competing vendors would offer a promising testing ground (e.g., IXI, Kantar's Worldpanel).

Imputing customer-level competitive marketing activities. Marketers have long recognized the importance of cross-effects in aggregate market response models. However, the effects of competitors' marketing efforts have been generally ignored in customer analytic models. While survey data could be collected, it may be difficult for customers to recall specific competitive offers and touchpoints. It might be feasible to ask customers for perceived relative levels of marketing efforts across competitors. Though imprecise, such information can be used to impute the general level of competitive efforts targeted toward a customer, which could highlight those customers who are potentially being targeted by competitors and whose share of wallet should be monitored for signs of a weakening relationship. In addition, it is possible for firms to monitor some of the online and direct marketing efforts by their competitors (e.g., Competiscan, Comperemedia) and incorporate these into their customer analytic models.

Competitive marketing efforts (e.g., which customer receives how much, and from whom?) may reflect what competitors know about customers that the firm does not. If this is the case, capturing competitive marketing efforts and incorporating them into a firm's customer analytic models may identify customers with more "winnable" growth potential. However, given the complexity of competitive efforts (e.g., high-dimensional, customer-specific, time-sensitive), care must be taken to account for the strategic nature of the marketing efforts by the firm and its competitors, and the possibility of signaling and counter response by competitors (Shin and

Sudhir 2010). When customer-level competitive intelligence is unavailable, data on aggregated competitive activities, such as those used for market response modeling, could offer a means to augment data for customer response modeling (e.g., modeling competitive efforts received by a customer as a function of aggregated competitive activities that vary across time and markets).

Quantifying the magnitude of competitive advantage from customer-level competitive intelligence. In the presence of competition, information asymmetries across firms with regard to customers' interactions will persist. Leveraging various data sources and statistical tools will not solve this problem completely. Compared with data on internal interactions, measures of external transactions and competitive marketing offers at the customer level are bound to be less accurate, comprehensive, or timely. This raises questions about the implications of firms' attempts to mitigate this information asymmetry. Chief among these questions is whether firms that acquire data on customers' external activities will have a competitive advantage over their counterparts and, if so, how large such an advantage would be and for how long it would persist.

In markets with multiple firms with significant market shares, the use of customer-level competitive intelligence may reveal a path to growth, be it through promoting products that customers have only purchased from competitors or refining the timing and level at which promotions are deployed in an effort to co-opt a customer journey that would have likely ended with a transaction with a competitor.

It is critical for research to consider the impact of acquiring customer-level competitive intelligence on the competitive equilibrium. Factors including the concentration of the market, heterogeneity in user preferences, and the extent to which customers prefer their data not to be shared by firms may determine whether customer-level competitive intelligence can support growth for those who acquire the information. If a firm can acquire and integrate customer-level competitive intelligence more efficiently than its competitors, such as by enticing its customers to actively share competitive efforts in exchange for benefits, it may have an opportunity to reap a competitive advantage over others. The firm's ability to sustain this advantage, though, depends on how quickly it can erect a moat around its newfound gains and stave off competitors' counter responses. The desire for customer-level competitive intelligence may also create an incentive for firms to pool information, whether collaborating directly or through a third-party entity, depending on the extent to which participants are collectively better or worse off.

Another potential outcome is that all firms acquire customer-level competitive intelligence, incurring the costs associated with acquiring and integrating the data but not having an opportunity to benefit from them, resulting in a Bertrand supertrap (Cabral and Vilas-Boas 2005). Theoretical models can evaluate the new competitive equilibrium, examining the effects on firms and customers, as well as identifying the factors that may moderate this equilibrium (e.g., Musalem and Joshi 2009; Shin and Sudhir 2010). Empirical research can also help address these questions through the use of single-source data in which all purchases made by a customer panel and their exposures to marketing are observed. Such data would enable researchers to build and evaluate customer response models under various data availability scenarios. Such analyses may offer firms guidance as to the potential growth they could gain from acquiring and incorporating customer-level competitive intelligence.

Customer Retention

Arguably the most financially relevant aspect of the customer equity framework is retention (Gupta, Lehmann, and Stuart 2004), putting churn management at the heart of CRM. Firms have traditionally relied on purchase and usage data to predict customer churn (e.g., Ascarza and Hardie 2013; Lemmens and Croux 2006). These internal data sources have been complemented with clickstream data, adding information on the customer's search process (Ascarza et al. 2018). While research on customer retention has focused on predicting churn, little attention has been given to mitigating churn, or such efforts have yielded little value (Ascarza et al. 2018).

This focus on churn prediction is yet another example of the streetlight effect. Once it has occurred, churn is observed, and it is relatively straightforward to leverage existing and emerging data sources to predict this outcome. Yet the goal for marketers should be to prevent churn that would have occurred otherwise. That is, mitigating churn is a counterfactual scenario that requires understanding *when* and *why* a customer may churn (e.g., Braun and Schweidel 2011) so that it can be prevented. We propose the use of two underutilized data sources in customer retention to focus on churn prevention: unstructured data and causal data that focus on proactive churn management.

Using Unstructured Data for Churn Management

If one wishes to move toward mitigating churn, marketers must explore why customers are unsatisfied and may be at risk of churn. One of the ways in which research can support firm growth through improved retention is by leveraging unstructured customer–firm interaction data to better inform who is likely to churn when and why. There are several challenges to doing so.

Challenges to leveraging unstructured customer–firm interaction data

Analyzing unstructured data. One of the key challenges associated with analyzing customer–firm communications is that the data from these interactions are unstructured, including textual data for chats, audio data for call center conversations, and video data for service encounters. Until recently, methods for automated unstructured data analysis have been limited. While we have seen increased interest in automated analyses of textual data (Berger et al. 2020), the development of methods for visual (e.g., Liu, Dzyabura, and Mizik 2020), audio, and video data (e.g., Liu et al. 2018) lags. Moreover, the analysis of unstructured data must be conducted over time and linked to customer behaviors to assess the relationship between customer–firm interactions and retention.

Overcoming privacy concerns. A contributing factor to the limited use of customer–firm interaction data is that such communications may include personal data such as credit card information or customers' names that are difficult to anonymize. As such, firms are often reluctant to share these data with service providers or researchers. Scrubbing such personal information requires analyzing the data with the same advanced unstructured analytic tools that the firm is trying to acquire to gain insights from customer–firm interactions.

Unstructured data can inform *what* customers do, *why* these customers may be unsatisfied, and *how* they can be retained. Firms spend billions of dollars on customer communications via advertising (Gordon et al. 2021), attending to nuances such as the font or background. Yet when the customer wants to communicate with the firm, despite being told that "this call may be recorded," the content of the interaction is often ignored. Apart from quality assurance, firms rarely investigate the conversation content systematically. While customers are notified of the recording for legal purposes, this seemingly benign statement may raise service expectations.

Opportunities and directions for future research in using unstructured interaction data. One of the biggest opportunities in managing churn is the development of automated methods to understand customer–firm interactions, be it through online chats, telephone calls, or direct encounters with service providers.[3]

Leveraging the dyadic nature of customer–firm interactions. From a methodological perspective, studying unstructured customer–firm interactions is novel for several reasons. In contrast to most common use of automated text analysis in marketing (e.g., analyzing UGC), which examines consumers' mass communication (Berger et al. 2020), understanding how customer–firm interactions affect retention requires analyzing the creator–receiver dyad. Considering a customer and the firm representative, analyses must identify the content creator and take conversational dynamics into account. Who says what and the preceding remarks from both sides of the dyad must be taken into account to recognize the context in which a comment is made. For instance, a comment about the price of service following a complaint about service quality may be handled differently than a comment about price following mentions of competitors.

In addition, based on the way in which customers interact with the firm, be it text, audio, or video, there is an opportunity to explore the stylistic similarity of the two agents (customer and firm representative). Previous research has demonstrated that a linguistic match between people can affect the interaction (e.g., Lemaire and Netzer 2020; Ludwig et al. 2013) due to aspects such as mimicry and homophily. The (mis)matching between the language, voice, or gestures of the firm representative and the customer may affect customer retention efforts. Future research could investigate how the stylistic similarity in text, audio, or imagery affects customer satisfaction and retention, paving the way for firms to train their representatives and best match representatives with customers based on their dispositions as revealed from prior interactions.

In the analysis of textual or voice conversations, the researcher is often interested in either exploring how the text or voice affects the receiver and/or what the text or voice reflects about the originator (Berger et al. 2020). Firms can improve their retention efforts by listening to their customers to identify what the conversation reflects about the customers' intentions. The weights that customers place on different aspects of service (e.g., price, quality, convenience) may be learned from the conversations that agents have over the course of the relationship. Such insights can be provided to agents prior to the start of a customer interaction so that they may be prepared with the language and potential offers that may be most effective at reducing the risk of customer churn. Alternatively, looking at the service representative, firms can investigate how the language that the service representative uses affects the customer's subsequent retention decisions. In addition to the impact of marketing offers that are provided based on what a customer has said, research could also investigate the effects of service representative empathy and listening. If such factors are found to affect customer retention decisions, this may call into question performance metrics related to the volume of calls that service representatives can field and the speed with which customer calls are dispatched.

Going beyond text—analyzing audio and video interactions. Customer–firm interactions may be multimodal, and the choice of communication mode may provide insight into the customer's mindset. Customers may choose to communicate through chatbots for the efficiency with which service can be delivered, and in the process of doing so only provide textual responses. Those deciding to contact the firm via phone for customer service provide both textual and audio information. The free-form nature of a dialog may be preferable for those who want to elaborate on their remarks or are seeking to express frustration. Other customers may prefer a person-to-person encounter, allowing them to express themselves through text, audio, and physical movements. These distinct modes of engagement pose a modeling challenge and an opportunity of how to best combine them to identify not only the topics of the conversation, but also the customer's state of mind (Matz and Netzer 2017) and the urgency with which the firm must respond to avert churn. Audio characteristics such as volume and tempo, as well as physical movements captured on video, may enable researchers to identify the interaction

[3] While we emphasize in this section the use of unstructured data for customer retention, the use of such data in all aspects of CRM, including customer acquisition and development, has been scant. Such data can also be beneficial in identifying leads for customer acquisition and opportunities for cross-sell, upsell, or share-of-wallet measurement.

characteristics that permit the earliest indication that a customer is at risk and for what reason. Such data may be more informative than the text transcript of the interaction in predicting the possibility of churn and understanding its possible causes.

Audio characteristics such as tone, pitch, and amplitude can be recorded and may reflect the customer's (and agent's) current state of mind during a customer–firm interaction. Such tools are sometimes used by call center management companies to identify the customer's mood from the pitch of their voice and, if necessary, escalate calls to more experienced agents. It has also been used to evaluate the impact of affective states during conference calls on firm performance (Mayew and Venkatachalam 2012). Investigating the content of past customer–firm interactions rather than simply focusing on the current interaction could capture changes in the relationship and help identify the customers who are most likely to churn. Deep learning methods have also been developed to analyze emotions based on audio content (e.g., Papakostas et al. 2017), which can be used to assess a customer's mindset across a sequence of interactions so as to spot changes in how customers communicate with the firm that may signal changes in the underlying relationship.

Identifying the reasons to churn. The content of customer–firm interactions can reveal not only whether a customer is likely to churn (e.g., mentioning the name of a competitor during a call center conversation), but also the underlying cause of churn. Interpreting statements such as "The competitor offered me a lower price" or "I don't get good reception in my basement" identifies the cause of churn, enabling firms to move from predicting to managing churn. In some cases, churn is outside the control of the firm (e.g., geographic relocation, death). When churn is preventable, understanding the cause can guide the firm's response (Braun and Schweidel 2011). A customer complaining about the price may be "saved" by a discount, whereas churning due to product features may be mitigated by a product upgrade.

Beyond the presence of specific keywords, other aspects of a dialog may be informative about the tendency to churn and its reasons. The use of pronouns by customer service agents, for example, have been found to affect customer satisfaction, with first-person pronouns being preferred to third-person pronouns (Packard, Moore, and Mcferran 2018). An analysis of customer churn based on the agents with whom customers have interacted may support such a pattern, indicating that training should be sufficiently detailed to coach agents on their linguistic choices. The pronoun choices by customers can also be examined to assess whether their usage patterns are indicative of churn in the near future. In addition to pronouns, other aspects of a customer's linguistic style can provide meaningful information about the individual, such as their traits or their emotional state during the interaction (Pennebaker 2011). Identifying the customer's emotional state can be useful in not only predicting churn, but also in establishing the timing and possible delivery of messages to prevent it.

Using biometric data in service encounters. Just as biometric data can inform acquisition efforts, it can also be gathered during customer–firm interactions and linked to the risk of churn. Research could focus on developing techniques to analyze real-time biometric data to inform how likely a customer is to churn, as well as to provide feedback to a service associate to mitigate churn. As discussed with regard to customer acquisition, research that applies these tactics to evaluate offline service encounters would be invaluable in shedding light on the drivers of churn.

Just as the audio of customer–firm interactions reflects a customer's emotional state, biometric measures can be extracted from videos of service encounters to analyze what contributed to them being a success or failure. Facial expressions and eye tracking may reveal a level of satisfaction with a given interaction that relates to a customer's retention decision (e.g., Bolton 1998). Future research could embark on the identification of signals from video footage that are indicative of future churn. Once such patterns have been identified, which requires establishing a link between what is being conveyed by an associate and the ensuing physiological response of the customer, tools can be developed to train service associates by identifying key moments in an encounter at which the customer responded (un)favorably. Research could also use biometric data to provide associates with real-time guidance on the information that should be shared with customers or when they would be better served by moving on to another customer.

Harnessing Causal Data for Proactive Retention

Much of firms' retention efforts are reactive as opposed to proactive (Ascarza et al. 2018), with firms waiting for the customer to exhibit indicators of impending churn and then offering incentives to stay, or even investing in win-back efforts after the customer churn. Such efforts often come up short because the customer has already made up their mind. If firms could identify customers who are on a path that is likely to result in churn, they may be able to deploy less costly efforts to change the customer's relationship trajectory.

Obtaining insights into customers' impending churn and the effectiveness of the firm's retention efforts can help firms move from reactive to proactive retention management (De Matos et al. 2018). Academic research on the effectiveness of such proactive campaigns is quite sparse (Ascarza 2018). Two key elements are needed to establish such proactive campaigns: (1) Which customers are at risk of churning? and (2) What are the effects of different proactive campaigns or messages on the risk of churning of different customers? While the first question has been heavily researched, the second question remains largely underexplored.

Challenges in harnessing causal data for proactive retention. Evaluating retention campaigns inherently requires a counterfactual analysis to assess their potential impact, resulting in several challenges that marketers must confront.

Accounting for the firm's targeting decisions with field data. As with other forms of field data, and particularly with respect to firms' targeted actions, customers' exposures and responses to retention efforts do not occur at random. Firms may target specific customers or groups of customers with retention efforts based on their past behaviors. Moreover, the content of these targeted campaigns is often nonrandom. Thus, efforts to isolate the effects of retention campaigns from field data require the ability to identify the target selection process or comparable groups of customers who were not targeted to serve as "controls."

Just as firms may be strategic in their deployment of retention efforts, so too are customers in their responses. Taking advantage of the fact that firms often provide incentives to dissatisfied customers to stay, strategic customers may complain to elicit such offers despite not having a grievance with the firm. Moreover, firm responses to complaints can increase customer expectations, resulting in more complaints (Ma, Sun, and Kekre 2015). As such strategic customer behavior is likely to be more pronounced in response to reactive rather than proactive campaigns, focusing on proactive retention may help mitigate such behavior.

Measuring responses over an appropriate time horizon. In contrast to responses to digital marketing efforts that can often be observed from click-through rates, evaluating retention campaigns requires a longer time horizon. One of the challenges this raises is which marketing efforts affect retention decisions throughout a customer's relationship with the firm. While the most recent customer touchpoint with the firm will affect retention decisions, so too may prior touchpoints, akin to the impact of earlier marketing activities in attribution models (e.g., Li and Kannan 2014). A customer's touchpoint history with the firm must be considered to evaluate the impact of marketing interventions on retention. An additional challenge to the extended time frame over which marketing interventions must be evaluated to assess their impact on retention is that other factors may affect customers' retention decisions during that period. Among these is the potential for spillover via WOM, which can result in customers who were not actually exposed to retention efforts becoming aware of them.

A final consideration related to the long-time horizon over which retention decisions occur is estimating the value associated with a retained customer. That is, suppose that we had a means of quantifying the incremental effect of a proactive retention campaign on a given customer's annual renewal decision. It would be shortsighted to only credit an amount equal to the expected increase in revenue attributable to the campaign for a single year. While the increase in residual lifetime value might be a theoretically sound measure, this would attribute credit for revenue that has not yet been realized.

Opportunities and future research directions in harnessing causal data for retention

Identifying naturally occurring variation in service levels. One way in which counterfactual scenarios can be evaluated is by gathering data that arise with naturally occurring variation. Such variation may create conditions akin to experiments that allow marketers to rule out other factors that might affect retention decisions.

Though identifying exogenous shocks may be difficult, we provide some illustrations of where they may occur. For service providers, exogenous shocks such as unanticipated service interruptions may affect certain customers. For example, poor reception due to weather conditions may affect only an isolated geographic area. An unexpected software glitch may cause a system outage that only affects those customers who tried to access the system during a narrow period of time. Introducing the service in one market may overwhelm the service center and have unintended consequences on the service in another unrelated geographical area. Researchers may examine the behavior of individuals before and after the service interruption, identifying those who were recipients of proactive retention efforts. As the service interruption may have affected the strength of customer relationships with the firm, we would anticipate that those who received the proactive marketing intervention would be less likely to churn following the service interruption, compared with those who did not receive such marketing efforts. The difference between these two groups offers a means of quantifying the extent to which proactive retention campaigns can mitigate the effects of service failures. Such quasi-experiments, or different matching algorithms through which customers who were "treated" with proactive retention efforts are matched with those who were untreated but have similar characteristics, may offer a means of controlling for nonrandom behavior by customers and firms.

Leveraging field experiments for proactive retention campaign design. An alternative to searching for naturally occurring variation is to undertake field experiments that have been designed to evaluate the long-term effects of retention campaigns. Field experiments have been used extensively in the domains of customer acquisition and response to marketing actions such as display advertising, website design, and catalog mailings (Feit and Berman 2019).

Despite their success in assessing the effectiveness of other aspects of marketing, the use of field experiments for the purpose of retention has been quite limited (for exceptions, see Ascarza [2018] and Lemmens and Gupta [2020]). Field experiments have been demonstrated to be useful in focusing the attention of management on customers who are likely to respond positively to retention efforts as opposed to those who are merely at risk of churning (Ascarza 2018). There are considerable opportunities to use field experiments to inform which retention efforts are likely to be useful for which customers and at what times throughout the relationship. Future research that combines field experiments with analysis of the reasons to churn can then provide firms with a means of matching the marketing effort that is likely to be most effective with each customer. For example, some services like online games or apps have fairly high churn rates within the first few days of usage. Tracking user behavior in these first few experiences

Table 1. Summary of Future Research Directions.

Focus Area	Opportunity	Future Research Directions
Customer acquisition	Incorporating biometric data	• Identifying biometric responses throughout the customer journey • Using biometrics to measure the effectiveness of acquisition efforts • Optimizing marketing messages in real time by using biometric data
	Incorporating social network	• Redefining the value of prospects for acquisition by incorporating their impact on others in their social network • Supplementing existing third-party profile data with social network data to improve targeting • Incorporating social influence into effort allocation and compensation models in B2B settings
Customer development	Leveraging trend data	• Identifying trends by synthesizing across data sources and domains • Conducting early long-range forecasting to separate meaningful trends from fads • Identifying trendsetting markets or segments for leading indicators of emergence, propagation and decline of trends
	Incorporating competitive intelligence	• Imputing size and share of wallet using multiple external data sources such as surveys, purchase panels, third party aggregators, clickstreams, social media, wearables, and mobile locations • Projecting the evolution of both size and share of wallet • Imputing customer-level competitive marketing activities • Quantifying the value of customer-level competitive intelligence
Customer retention	Using unstructured data	• Analyzing the dyadic nature of customer–firm interactions between the customer and the service provider. What does the conversation reflect about the customer and how can the service provider affect the customer? • Exploring customers' state of mind and response to service interactions from audio and video data • Using textual data to identify reasons to churn and their implications for proactive churn management • Using biometric data to improve service encounters
	Harnessing causal data for proactive retention	• Leveraging naturally occurring variation such as exogenous shocks or service failures to examine the causal impact of retention efforts • Leveraging field experiments and measures such as real-time satisfaction to match retention efforts with customers • Exploring the social effects of proactive retention efforts
Managerial challenges		• Quantifying the incremental value of marketing data and applications • Recognizing the potential ethical and legal costs of marketing data and their impact on customers and firms • Prioritizing the data and application mix for firms along the analytics maturity curve • Understanding the sustainability of data-driven growth

and understanding which aspects of the product grab customers' attention may help designing targeted proactive retention campaigns that better meet customer needs. Field experiments can also get at the importance of the timing with which marketing efforts are deployed.

In designing and deploying retention efforts, incorporating individual-level measures such as customer satisfaction is a promising direction to pursue. Mobile, digital and in-store electronic devices enable firms to track satisfaction in real time, not just with respect to the product but also with respect to other aspects of the experience. Such data allow firms to experiment with retention efforts that match aspects of the experience with time periods of low satisfaction (e.g., Nevskaya and Albuquerque 2019). We encourage future research to combine the rich work on customer satisfaction with the causal effects of proactive retention efforts.

Measuring social effects of proactive retention campaigns. Just as social ties may affect customer acquisition, they may be instrumental in retention decisions, as a focal customer with social ties who have a high churn rate may also have a high risk of churning (De Matos et al. 2018). Because it is difficult to separate social effects from homophily in historical data (Manski 1993), data derived from social field experiments can isolate the causal social effects of churn (e.g., Ascarza et al. 2017). One potential avenue to explore is the derivation of social scores for customers with regard to their influence on the retention decisions of others. Such measures may be derived based on the focal customer's ties to other customers and the focal customer's role in being a bridge between different communities of customers. We expect such social effects to be particularly important for services that benefit from network externality such as online games, cloud collaboration platforms and telecommunication services. By seeding firm-generated proactive retention messages with different customers based on their location in the customer network, researchers can estimate the extent to which social influence may be a tool they can leverage to amplify the impact of retention efforts.

Overall, we believe that causal data, whether through naturally occurring variation or field experiments, can help focus research and CRM practice on proactive churn management, moving beyond mere prediction. In doing so, the field of CRM can move beyond descriptive and predictive research, and toward more theory-driven causal inference and prescriptive research.

General Discussion

In response to a top MSI research priority—capturing information to fuel growth—we focused on growth via the customer equity framework and discussed six data areas where we see substantial opportunities. Table 1 summarizes the promising research directions we identified.

Given the breadth of the topic, it would be impossible to enumerate all the paths through which data can be turned into growth opportunities. At a fundamental level, the entirety of marketing analytics is focused on deriving information from data. Our focus is on illustrating the connection between marketing data and the three drivers of customer equity as organic growth avenues—acquiring new customers, developing existing customers, and retaining them. In doing so, we highlight six areas of data and applications that may have been overlooked due to the streetlight effect but can be potent for firms pursuing organic growth.

We hope our (admittedly selective) discussions illustrate the opportunity costs that a data availability bias creates when it comes to searching for data-driven growth. While marketing technology and analytics have advanced, they have also exacerbated the risk of falling victim to the streetlight effect as newer, bigger, and richer data are constantly thrust on marketers, often without a clear roadmap to drive growth. Despite the siren song of novel data sources, marketers must stay apprised of the ever-expanding data landscape and become familiar with the potential use cases of a wide range of data sources. Indeed, companies have emerged to create more transparency and trust in the data marketplace (e.g., Datarade, AlternativeData.org, AWS Data Exchange, ProgrammableWeb), offering to connect curated data providers with data buyers.

Marketers must stop thinking about the use of data solely in terms of how the available information can be analyzed. Rather, they need to consider the use of data as a component of a strategy problem. That is, how can existing and nascent data resources be brought into alignment with the firm's growth strategy? We believe this seemingly subtle difference in managerial orientation can make a substantial difference in turning data opportunities into growth opportunities, reducing preoccupation with short-term operational goals and tactical problems.

While we have focused on opportunities associated with leveraging select data sources, resolving the disconnect between data and firm growth requires addressing not only data and analytics challenges, but also managerial challenges. Deshpandé and Zaltman (1982) discuss drivers of data and analysis adoption by management, finding that less centralized and less formalized organizations are more likely to make use of data analysis. The misalignment between data mix and growth strategy may stem from the analytics teams and marketing decision makers being at arm's length in a centralized organization. They also find that the involvement of management in data collection and research is crucial for adoption of data-driven decision making. Surprising results of data analysis were found to limit the use of data analysis in decision making, risking a confirmation bias of leveraging data only when it leads to the preconceived directional results. Of course, researchers need to further examine how robust these organizational phenomena are and identify the underlying processes. To conclude, we discuss a few broader areas that can affect the adoption of data-driven growth by firms and in which our specific observations can be couched.

Quantifying the incremental value of marketing data and applications for growth. Marketers must evaluate the incremental benefit of each data resource and make cost–benefit trade-offs in determining whether to generate it internally, acquire it, or forgo it, as well as which application are best fitted for each data source. The challenge in determining this trade-off is exacerbated when firms are faced with an explosion of alternatives. The literature offers little guidance on how marketers can build a data and application portfolio suitable for their growth strategies and budgets. For targeting and personalization, how can one quantify the marginal value versus cost of each additional piece of individual data? For monitoring brand equity, what is the optimal mix of data collected through tracking surveys and social listening? For media measurement and planning, what is the optimal mix of individual data for multitouch attribution modeling and aggregate data for marketing-mix modeling? For idea generation, what is the optimal mix of data collected through qualitative methods (e.g., focus groups, depth interviews, ethnographic observations) versus data gathered through machine learning based on UGC?

We have little evidence that quantifies the ROI firms can generate from investing in different data sources and applications (Berman and Israeli 2021). Such research would offer tremendous value to marketers as they aim to optimize their data and application mix. One approach that firms may consider is conducting field experiments (akin to the one discussed in the "Customer Retention" section) in which customers are targeted with different treatments based on different sources of data, or with different levels of data usage (e.g., with or without augmenting internal data with customer profile data from third-party providers). The firm would then be able to evaluate how its marketing decisions and customer responses would differ if a particular data source were incorporated into its decision making. The incremental performance can then be weighed against the incremental costs of data acquisition and analysis. Similarly, data suppliers can experiment with the data they provide to different organizations or use quasi–field experiments to build case studies that demonstrate the ROI of using their data.

Recognizing the full cost of marketing data for growth. Marketers should also consider the potential ethical and legal costs of the data they employ. For example, investing in cross-device identity resolution can help connect customer behavior from diverse devices to create a more holistic profile that informs targeting decisions and personalization. However, such rich individually identifiable data may have hidden costs that are difficult to quantify, such as changes in customer behavior due to privacy concerns, heightened risks of data breaches, and increasing costs of regulatory compliance with General Data Protection Regulation, California Consumer Privacy Act, and other emerging legislations and regulations (Cui et al. 2021).

With the trend toward giving customers more control over how and by whom their data are collected, there can be costly implications for data-driven growth. What if a portion of customers decide they do not want to share their data with a firm? This could lead to imbalanced data across customers, inadvertently creating data-driven discrimination that can have unintended consequences. Future research should aim to develop methods by which available data can be used to derive inferences about those who elect to not share their data with firms. Using data from third-party aggregators (e.g., Acxiom) and public social media posts, for instance, could allow some of the missing individual data to be imputed. In doing so, firms can make use of the data that consumers have purposefully shared with the public while respecting their privacy.

We are also seeing more regulations of specific types of data (e.g., the ban on sales of location data from cell phone companies in New York City) and calls to not use information for specific purposes (e.g., banning facial recognition use by police). How do marketers engage customers and regulators to convey the value of marketing data? Firms must be transparent with customers and disclose how customer-provided data will be used, and what value customers will receive in exchange for such data. Delivering on this promise will entail firms going on a data "diet," only requesting essential data from customers. Such transparency and a reduction in the stockpiling of customer data may assuage privacy concerns. It may also be prudent for marketers and future researchers to explore privacy-friendly approaches to data-driven growth. For example, the adoption and reporting of K-anonymity (Sweeney 2002), in which an individual in a data set cannot be distinguished from at least K-1 other individuals who are also in the data set, could allay concerns about the invasiveness of marketing data collection and analytics.

Moving along the marketing data and analytics maturity curve. While the academic literature favors research using data from novel and emerging sources, firms differ in terms of their state of data and analytics maturity (DalleMule and Davenport 2017). Many firms are still in the early stages of developing a marketing data infrastructure. Research should identify the optimal path forward for firms at different stages of data-driven decision-making readiness. Should firms relatively new to data-driven decisions first focus on building a robust CRM system, followed by a system to track marketing expenses and then collecting the "digital exhaust" (e.g., search behavior, online clickstream, social media posts)? When is the right time to think about brand tracking or competitive intelligence?

Firms also need guidance on the appropriate balance between investments in data generation/acquisition and data applications. Firms in the early stages of the analytic maturity curve may focus on building application capabilities, which can help harvest the "low-hanging fruits" from existing data. As firms mature analytically, they may rebalance their investments to data generation/acquisition efforts that can yield a competitive edge. Similarly, should young firms primarily focus on data related to customer acquisition, and invest in data related to customer development and retention only in later stages? Without evidence-based guidance, firms that are less mature analytically will face difficulty in leveraging data for growth.

Turning marketing data mix and applications into a sustainable competitive advantage. A key managerial challenge about data-driven growth lies in the development of a sustainable competitive advantage when the same data sources and applications are available to all. For example, American Airlines pioneered yield management to reap large but short-lived benefits, as yield management is now the norm among modern airlines. Marketers need to outthink and out-execute their competition in ensuring the uniqueness of their data mix and uses, lest the growth opportunities they expected be short-lived due to the commoditization of data.

This challenge raises many important research questions. For example, what types of information are harder for competitors to match (e.g., combining proprietary internal data with external data to derive insights)? What types of analytical tools are harder for competitors to duplicate, thus allowing firms to see what everybody else sees, but think what nobody else does (e.g., proprietary algorithms for extracting insights from unstructured data)? To what extent can those unique data and analytical tools sustain a competitive advantage? Having answers to these questions can illuminate our thinking about the strategic challenge of turning marketing data and analytics into a sustainable driver of superior growth.

Building on the work of Hagiu and Wright (2020), we see several key aspects that can help determine when marketing data create sustainable competitive advantages. First, is the data proprietary? This condition certainly favors internal over external data. We believe it also favors unstructured over structured data because the former could potentially allow far more degrees of freedom in how it is used for data-enabled learning.

Second, how long do data remain relevant? Short-shelf-life data make it difficult to build sustainable data-driven competitive advantages. Marketers need to be keenly aware of the shelf life of their data, models, and algorithms. More research is needed in devising scientific methods for determining how often data and applications need to be refreshed. Using forward-looking data and analytics such as the trendspotting analysis we discussed previously can be useful to mitigate that risk.

Third, how much does the data application benefit the customer? Applications such as data-enabled personalization can create more defensible moats than, say, targeting, because the former offers more value to the customer. However, while data-enabled personalization increases the switching cost for that one customer, it does not provide an advantage in competing for new customers. More sustainable data-enabled learning would allow information "spillovers" from existing customers to new customers.

Fourth, how fast can the insights from data be incorporated into offerings? The faster the firm can bring data-enabled learnings to the market while managing data privacy concerns (Kalaignanam et al. 2021), the harder it is for competitors to catch up. Research should guide marketers to balance insights generated from analytics projects and the amount of time and resources it takes to run them. The "test and learn" framework is useful in expediting the implementation of data-driven insights. Furthermore, how difficult is it to imitate product improvements that are based on customer data? The key factor affecting firms' ability to overcome this challenge is whether the data-enabled improvements are hidden or deeply embedded in a complex production process, making them hard to replicate.

Conclusion

Whereas the data available to marketers are vast and proliferating, too often academics and practitioners view the acquisition of data and subsequent analyses as the end goal. Data that are most accessible or lend themselves to easily interpretable analyses may be preferred. Such a streetlight effect in the utilization of marketing data and applications can result in missed growth opportunities if they require data that are less accessible or harder to interpret.

Rooted in the customer equity framework, we encourage viewing marketing data and the applications they enable as a component that supports firm growth. Beyond the identification of relevant data sources and the ability to implement the appropriate analyses, this perspective requires that firms be mindful of the managerial challenges they may face in leveraging data for firm growth. We hope our discussions illustrate several data-driven growth opportunities that may have gone overlooked and serve as a call to action for researchers and practitioners to better align marketing data and analytics with firm growth.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Oded Netzer https://orcid.org/0000-0002-0099-8128
David A. Schweidel https://orcid.org/0000-0003-2665-3272

References

Anderson, Eric, Song Lin, Duncan Simester, and Catherine Tucker (2015), "Harbingers of Failure," *Journal of Marketing Research*, 52 (5), 580–92.

Ascarza, Eva (2018), "Retention Futility: Targeting High-Risk Customers Might Be ineffective," *Journal of Marketing Research*, 55 (1), 80–98.

Ascarza, Eva, Peter Ebbes, Oded Netzer, and Matthew Danielson (2017), "Beyond the Target Customer: Social Effects of Customer Relationship Management Campaigns," *Journal of Marketing Research*, 54 (3), 347–63.

Ascarza, Eva and Bruce G.S. Hardie (2013), "A Joint Model of Usage and Churn in Contractual Settings," *Marketing Science*, 32 (4), 570–90.

Ascarza, Eva, Scott A. Neslin, Oded Netzer, Zachery Anderson, Peter S. Fader, Sunil Gupta, et al. (2018), "In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions," *Customer Needs and Solutions*, 5 (1/2), 65–81.

Bell, David R. and Sangyoung Song (2007), "Neighborhood Effects and Trial on the Internet: Evidence from Online Grocery Retailing," *Quantitative Marketing and Economics*, 5 (4), 361–400.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer, and David A. Schweidel (2020), "Uniting the Tribes: Using Text for Marketing Insight," *Journal of Marketing*, 84 (1), 1–25.

Berman, Ron and Ayelet Israeli (2021), "The Added Value of Data-Analytics: Evidence from Online Retailers," working paper.

Blattberg, Robert C. and John Deighton (1996), "Manage Marketing by the Customer Equity Test," *Harvard Business Review*, 74 (4), 136–44.

Bolton, Ruth N. (1998), "A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction," *Marketing Science*, 17 (1), 45–65.

Boulding, William, Richard Staelin, Michael Ehret, and Wesley J. Johnston (2005), "A Customer Relationship Management

Roadmap: What Is Known, Potential Pitfalls, and Where to Go," *Journal of Marketing*, 69 (4), 155–66.

Braun, Michael and André Bonfrer (2011), "Scalable Inference of Customer Similarities from Interactions Data Using Dirichlet Processes," *Marketing Science*, 30 (3), 513–31.

Braun, Michael and David A. Schweidel (2011), "Modeling Customer Lifetimes with Multiple Causes of Churn," *Marketing Science*, 30 (5), 881–902.

Bulling, Andreas and Michel Wedel (2019), "Pervasive Eye-Tracking for Real-World Consumer Behavior Analysis," in *Handbook of Process Tracing Methods*, Michael Schulte-Mecklenbeck, Anton Kuehberger and Joseph G. Johnson, eds. Abingdon, UK: Routledge, 27–44.

Cabral, Luís and J. Miguel Villas-Boas (2005), "Bertrand Supertraps," *Management Science*, 51 (4), 599–613.

Chandrasekaran, Deepa and Gerard J. Tellis (2007), "A Critical Review of Marketing Research on Diffusion of New Products," *Review of Marketing Research*, 3 (1), 39–80.

Chen, Xinlei, Yuxin Chen, and Ping Xiao (2013), "The Impact of Sampling and Network Topology on the Estimation of Social Intercorrelations," *Journal of Marketing Research*, 50 (1), 95–110.

Chen, Yuxin and Joel H. Steckel (2012), "Modeling Credit Card Share of Wallet: Solving the Incomplete Information Problem," *Journal of Marketing Research*, 49 (5), 655–69.

Cheng, Andria (2019), "Healthcare May Eventually Become a Bigger Business for Best Buy Than Selling Electronics," *Forbes* (September 24), https://www.Forbes.Com/Sites/Andriacheng/2019/09/24/This-Business-May-Eventually-Be-Bigger-For-Best-Buy-Than-Selling-Electronics/.

Cui, Toni, Anindya Ghose, Hanna Halaburda, Raghuram Iyengar, Koen Pauwels, S. Sriram, et al. (2021), "Informational Challenges in Omnichannel Marketing: Remedies and Future Research," *Journal of Marketing*, 85 (1), 103–20.

Dallemule, L. and T.H. Davenport (2017), "What's Your Data Strategy?" *Harvard Business Review*, 95 (3), 112–21.

Davenport, Thomas and Jeanne Harris (2017), *Competing on Analytics: Updated, with a New Introduction: The New Science of Winning*. Boston: Harvard Business Press.

De, Matos, Godinho Miguel, Ferreira Pedro, and Belo Rodrigo (2018), "Target the Ego or Target the Group: Evidence from a Randomized Experiment in Proactive Churn Management," *Marketing Science*, 37 (5), 793–811.

Deshpandé, Rohit and Gerald Zaltman (1982), "Factors Affecting the Use of Market Research Information: A Path Analysis," *Journal of Marketing Research*, 19 (1), 14–31.

Dew, Ryan and Asim Ansari (2018), "Bayesian Nonparametric Customer Base Analysis with Model-Based Visualizations," *Marketing Science*, 37 (2), 216–35.

Dover, Yaniv, Jacob Goldenberg, and Daniel Shapira (2012), "Network Traces on Penetration: Uncovering Degree Distribution from Adoption Data," *Marketing Science*, 31 (4), 689–712.

Drucker, Peter F. (1954), *The Practice of Management*. New York: Harper & Row Publishers, Inc.

Du, Rex Y., Ye Hu, and Sina Damangir (2015), "Leveraging Trends in Online Searches for Product Features in Market Response Modeling," *Journal of Marketing*, 79 (1), 29–43.

Du, Rex Y. and Wagner A. Kamakura (2006), "Household Life Cycles and Lifestyles in the United States," *Journal of Marketing Research*, 43 (1), 121–32.

Du, Rex Y. and Wagner A. Kamakura (2012), "Quantitative Trendspotting," *Journal of Marketing Research*, 49 (4), 514–36.

Du, Rex Y., Wagner A. Kamakura, and Carl F. Mela (2007), "Size and Share of Customer Wallet," *Journal of Marketing*, 71 (2), 94–113.

Eastlack, Joseph O., Jr., and Ambar G. Rao (1989), "Advertising Experiments at the Campbell Soup Company," *Marketing Science*, 8 (1), 57–71.

eMarketer (2020), "US Retail Sales to Drop More Than 10% in 2020," (June 7), https://www.emarketer.com/content/us-retail-sales-drop-more-than-10-2020.

Ester, Mike and Tripp Mickle (2015), "PepsiCo to Drop Aspartame from Diet Pepsi: Consumer Backlash, Slumping Sales Prompt Beverage Giant to Switch Artificial Sweeteners," *The Wall Street Journal*, (April 24), https://www.wsj.com/articles/pepsico-to-replace-aspartame-with-sucralose-in-diet-pepsi-in-u-s-1429 885941#:∼:text=PepsiCo%20Inc.,controversial%20but%20 still%20artificial%20 sweetener.

Favaretto, Rodolfo Migon, Paulo Knob, Soraia Raupp Musse, Felipe Vilanova, and Ângelo Brandelli Costa (2019), "Detecting Personality and Emotion Traits in Crowds from Video Sequences," *Machine Vision and Applications*, 30 (5), 999–1012.

Feit, Elea Mcdonnell and Ron Berman (2019), "Test & Roll: Profit-Maximizing A/B Tests," *Marketing Science*, 38 (6), 1038–58.

Golder, Peter N., Rachel Shacham, and Debanjan Mitra (2009), "Findings—Innovations' Origins: When, by Whom, and How Are Radical Innovations Developed?" *Marketing Science*, 28 (1), 166–79.

Gordon, Brett R., Kinshuk Jerath, Zsolt Katona, Sridhar Narayanan, Jiwoong Shin, and Kenneth C. Wilbur (2021), "Inefficiencies in Digital Advertising Markets," *Journal of Marketing*, 85 (1), 7–25.

Guadagni, Peter M. and John D.C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (3), 203–38.

Gupta, Sunil, Donald R. Lehmann, and Jennifer Ames Stuart (2004), "Valuing Customers," *Journal of Marketing Research*, 41 (1), 7–18.

Gupta, Sunil and Carl F. Mela (2008), "What Is a Free Customer Worth? Armchair Calculations of Nonpaying Customers' Value Can Lead to Flawed Strategies," *Harvard Business Review*, 86 (11), 102–09.

Hagiu, A. and Julian Wright (2020), "When Data Creates Competitive Advantage," *Harvard Business Review*, 98 (1), 94–101.

Hanssens, Dominique M. and Koen H. Pauwels (2016), "Demonstrating the Value of Marketing," *Journal of Marketing*, 80 (6), 173–90.

Heist, Greg and Sarah Tarraf (2016), "Trend Analytics: A Data-Driven Path to Foresight," *Marketing Insights*, Spring, 18–19.

Hersh, Eitan D. and Brian F. Schaffner (2013), "Targeted Campaign Appeals and the Value of Ambiguity," *Journal of Politics*, 75 (2), 520–34.

Ho, Teck-Hua, Shan Li, So-Eun Park, and Zuo-Jun Max Shen (2012), “Customer Influence Value and Purchase Acceleration in New Product Diffusion,” *Marketing Science*, 31 (2), 236–56.

Hoffman, Donna L. and Thomas P. Novak (1996), “Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations,” *Journal of Marketing*, 60 (3), 50–68.

Iyengar, Raghuram, Christophe Van Den Bulte, and Thomas W. Valente (2011), “Opinion Leadership and Social Contagion in New Product Diffusion,” *Marketing Science*, 30 (2), 195–212.

Kalaignanam, Kartik, Kapil R. Tuli, Tarun Kushwaha, Leonard Lee, and David Gal (2021), “Marketing Agility: The Concept, Antecedents, and a Research Agenda,” *Journal of Marketing*, 85 (1), 35–58.

Keiningham, Timothy L., Lerzan Aksoy, Alexander Buoye, and Bruce Cooil (2011), “Customer Loyalty Isn’t Enough. Grow Your Share of Wallet,” *Harvard Business Review*, 89 (10), 29–31.

Keiningham, Timothy L., Alexander Buoye, and Joan Ball (2015), “Competitive Context Is Everything: Moving from Absolute to Relative Metrics,” *Global Economics and Management Review*, 20 (2), 18–25.

Kelso, Alicia (2018), “Diet Pepsi Is Bringing Aspartame Back, Again,” Food Dive (February 22), https://www.fooddive.com/news/diet-pepsi-is-bringing-aspartame-back-again/517618/#:∼:text=Dive%20Brief%3A,but%20only%20via%20e%2 Dcommerce.

Kumar, V. (2018), “A Theory of Customer Valuation: Concepts, Metrics, Strategy, and Implementation,” *Journal of Marketing*, 82 (1), 1–19.

Kumar, V., J. Andrew Petersen, and Robert P. Leone (2010), “Driving Profitability by Encouraging Customer Referrals: Who, When, and How,” *Journal of Marketing*, 74 (5), 1–17.

Lambrecht, Anja and Catherine Tucker (2019), “Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads,” *Management Science*, 65 (7), 2966–81.

Lemaire, Alain and Oded Netzer (2020), “Linguistic-Based Recommendation: The Role of Linguistic Match Between Users and Products,” working paper, Columbia University.

Lemmens, Aurélie and Christophe Croux (2006), “Bagging and Boosting Classification Trees to Predict Churn,” *Journal of Marketing Research*, 43 (2), 276–86.

Lemmens, Aurélie and Sunil Gupta (2020), “Managing Churn to Maximize Profits,” *Marketing Science*, 39 (5), 956–73.

Lemon, Katherine N. and Peter C. Verhoef (2016), “Understanding Customer Experience Throughout the Customer Journey,” *Journal of Marketing*, 80 (6), 69–96.

Levinovitz, Alan (2015), *The Gluten Lie: And Other Myths About What You Eat*. New York: Simon and Schuster.

Lewis, Michael (2006), “Customer Acquisition Promotions and Customer Asset Value,” *Journal of Marketing Research*, 43 (2), 195–203.

Li, Hongshuang (Alice) and P.K. Kannan (2014), “Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment,” *Journal of Marketing Research*, 51 (1), 40–56.

Liu, Liu, Daria Dzyabura, and Natalie Mizik (2020), “Visual Listening in: Extracting Brand Image Portrayed on Social Media,” *Marketing Science*, 39 (4), 669–86.

Liu, Xuan, Savannah Wei Shi, Thales Teixeira, and Michel Wedel (2018), “Video Content Marketing: The Making of Clips,” *Journal of Marketing*, 82 (4), 86–101.

Lodish, Leonard M. and Carl F. Mela (2007), “If Brands Are Built over Years, Why Are They Managed over Quarters?” *Harvard Business Review*, 85 (7/8), 104–12.

Ludwig, Stephan, Ko De Ruyter, Mike Friedman, Elisabeth C. Brüggen, Martin Wetzels, and Gerard Pfann (2013), “More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates,” *Journal of Marketing*, 77 (1), 87–103.

Ma, Liye, Baohong Sun, and Sunder Kekre (2015), “The Squeaky Wheel Gets the Grease—An Empirical Analysis of Customer Voice and Firm Intervention on Twitter,” *Marketing Science*, 34 (5), 627–45.

Manski, Charles F. (1993), “Identification of Endogenous Social Effects: The Reflection Problem,” *Review of Economic Studies*, 60 (3), 531–42.

Matz, Sandra C. and Oded Netzer (2017), “Using Big Data as a Window into Consumers’ Psychology,” *Current Opinion in Behavioral Sciences*, 18, 7–12.

Mayew, William J. and Mohan Venkatachalam (2012), “The Power of Voice: Managerial Affective States and Future Firm Performance,” *Journal of Finance*, 67 (1), 1–43.

McCarthy, Daniel M., Peter S. Fader, and Bruce G.S. Hardie (2017), “Valuing Subscription-Based Businesses Using Publicly Disclosed Customer Data,” *Journal of Marketing*, 81 (1), 17–35.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, Joseph P. Pickett, et al. (2011), “Quantitative Analysis of Culture Using Millions of Digitized Books,” *Science*, 331 (6014), 176–82.

Minsky, Laurence and Keith A. Quesenberry (2016), “How B2B Sales Can Benefit from Social Selling,” *Harvard Business Review* (November 8), https://hbr.org/2016/11/84-of-b2b-sales-start-with-a-referral-not-a-salesperson.

Mintel (2018), “US Gluten-Free Foods Market Report,” https://store.mintel.com/us-gluten-free-foods-market-report.

Mitra, Debanjan and Peter N. Golder (2002), “Whose Culture Matters? Near-Market Knowledge and Its Impact on Foreign Market Entry Timing,” *Journal of Marketing Research*, 39 (3), 350–65.

Moe, Wendy W. and David A. Schweidel (2012), “Online Product Opinions: Incidence, Evaluation, and Evolution,” *Marketing Science*, 31 (3), 372–86.

Moon, Sangkil, Wagner A. Kamakura, and Johannes Ledolter (2007), “Estimating Promotion Response When Competitive Promotions Are Unobservable,” *Journal of Marketing Research*, 44 (3), 503–15.

Musalem, Andrés and Yogesh V. Joshi (2009), “Research Note—How Much Should You Invest in Each Customer Relationship? A Competitive Strategic Approach,” *Marketing Science*, 28 (3), 555–65.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), “Mine Your Own Business: Market-Structure Surveillance Through Text Mining,” *Marketing Science*, 31 (3), 521–43.

Netzer, Oded, James M. Lattin, and Vikram Srinivasan (2008). "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science*, 27 (2), 185–204.

Neumann, Nico, Catherine E. Tucker, and Timothy Whitfield (2019), "Frontiers: How Effective Is Third-Party Consumer Profiling? Evidence from Field Studies," *Marketing Science*, 38 (6), 918–26.

Nevskaya, Yulia and Paulo Albuquerque (2019), "How Should Firms Manage Excessive Product Use? A Continuous-Time Demand Model to Test Reward Schedules, Notifications, and Time Limits," *Journal of Marketing Research*, 56 (3), 379–400.

Nocera, Chris (2020), "People with This Blood Type Were at Less Risk of Experiencing Severe COVID-19 Symptoms, Study Showed," ABC (July 11), https://abc13.com/Covid-19-Blood-Type-O-23andme/6310236/.

Ofek, Elie and Luc Wathieu (2010), "Are You Ignoring Trends That Could Shake Up Your Business?" *Harvard Business Review*, 88 (7/8), 124–31.

Packaged Facts (2016), *Gluten-Free Foods in the U.S.*, 6th ed, https://www.packagedfacts.com/Gluten-Free-Foods-10378213/.

Packard, Grant, Sarah G. Moore, and Brent Mcferran (2018), "(I'm) Happy to Help (You): The Impact of Personal Pronoun Use in Customer–Firm interactions," *Journal of Marketing Research*, 55 (4), 541–55.

Papakostas, Michalis, Evaggelos Spyrou, Theodoros Giannakopoulos, Giorgos Siantikos, Dimitrios Sgouropoulos, Phivos Mylonas, et al. (2017), "Deep Visual Attributes vs. Hand-Crafted Audio Features on Multidomain Speech Emotion Recognition," *Computation*, 5 (2), 26.

Pennebaker, James W. (2011), "The Secret Life of Pronouns," *New Scientist*, 211 (2828), 42–45.

Pieters, Rik and Michel Wedel (2004), "Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects," *Journal of Marketing*, 68 (2), 36–50.

Pozharliev, Rumen, Willem J.M.I. Verbeke, Jan W. Van Strien, and Richard P. Bagozzi (2015), "Merely Being with You Increases My Attention to Luxury Products: Using EEG To Understand Consumers' Emotional Experience with Luxury Branded Products," *Journal of Marketing Research*, 52 (4), 546–58.

Puntoni, Stefano, Rebecca Walker Reczek, Markus Giesler, and Simona Botti (2021), "Consumers and Artificial Intelligence: An Experiential Perspective," *Journal of Marketing*, 85 (1), 131–51.

Schoenmueller, Verena, Oded Netzer, and Florian Stahl (2020), "The Polarity of Online Reviews: Prevalence, Drivers and Implications," *Journal of Marketing Research*, 57 (5), 853–77.

Schweidel, David A., Eric T. Bradlow, and Peter S. Fader (2011), "Portfolio Dynamics for Customers of a Multiservice Provider," *Management Science*, 57 (3), 471–86.

Schweidel, David A. and Wendy W. Moe (2014), "Listening in on Social Media: A Joint Model of Sentiment and Venue Format Choice," *Journal of Marketing Research*, 51 (4), 387–402.

Shin, Jiwoong and K. Sudhir (2010), "A Customer Management Dilemma: When Is It Profitable to Reward One's Own Customers?" *Marketing Science*, 29 (4), 671–89.

Shugan, Steven M. and Debanjan Mitra (2009), "Metrics—When and Why Non-Averaging Statistics Work," *Management Science*, 55 (1), 4–15.

Shugan, Steven M. and Debanjan Mitra (2014), "A Theory for Market Growth or Decline," *Marketing Science*, 33 (1), 47–65.

Sweeney, Latanya (2002), "K-Anonymity: A Model for Protecting Privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10 (5), 557–70.

The CMO Survey (2020), "Highlights and Insights Report," (February), https://CMOsurvey.org/Results/February-2020/.

Tonietto, Gabriela N. and Alixandra Barasch (2020), "Generating Content Increases Enjoyment by Immersing Consumers and Accelerating Perceived Time," *Journal of Marketing* (published online September 10), DOI:10.1177/0022242920944388.

Trusov, Michael, Randolph E. Bucklin, and Koen Pauwels (2009), "Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing*, 73 (5), 90–102.

U.S. Food and Drug Administration (2018), "FDA Authorizes, with Special Controls, Direct-to-Consumer Tests that Reports Three Mutations in the BRCA Breast Cancer Genes," press release (March 6), https://www.fda.gov/news-events/press-announcements/fda-authorizes-special-controls-direct-consumer-test-reports-three-mutations-brca-breast-cancer.

Venkatraman, Vinod, Angelika Dimoka, Paul A. Pavlou, Khoi Vo, William Hampton, Bryan Bollinger, et al. (2015), "Predicting Advertising Success Beyond Traditional Measures: New Insights from Neurophysiological Methods and Market Response Modeling," *Journal of Marketing Research*, 52 (4), 436–52.

Villanueva, Julian, Shijin Yoo, and Dominique M. Hanssens (2008), "The Impact of Marketing-Induced Versus Word-of-Mouth Customer Acquisition on Customer Equity Growth," *Journal of Marketing Research*, 45 (1), 48–59.

Watts, Jameson (2018), "Trend Spotting: Using Text Analysis to Model Market Dynamics," *International Journal of Market Research*, 60 (4), 408–18.

Yao, Song and Carl F. Mela (2008), "Online Auction Demand," *Marketing Science*, 27 (5), 861–85.

Yoon, Carolyn, Richard Gonzalez, and James R. Bettman (2009), "Using fMRI to Inform Marketing Research: Challenges and Opportunities," *Journal of Marketing Research*, 46 (1), 17–19.

Zhong, Ning and David A. Schweidel (2020), "Capturing Changes in Social Media Content: A Multiple Latent Changepoint Topic Model," *Marketing Science*, 39 (4), 827–46.

Zubcsek, Peter Pal, Zsolt Katona, and Miklos Sarvary (2017), "Predicting Mobile Advertising Response Using Consumer Colocation Networks," *Journal of Marketing*, 81 (4), 109–26.