Mining Consumer Minds: Downstream Consequences of Host Motivations for Home-Sharing Platforms

JAEYEON (JAE) CHUNG
GITA VENKATARAMANI JOHAR
YANYAN LI
ODED NETZER
MATTHEW PEARSON

This research sheds light on consumer motivations for participating in the sharing economy and examines downstream consequences of the uncovered motivations. We use text-mining techniques to extract Airbnb hosts' motivations from their responses to the question "why did you start hosting." We find that hosts are driven not only by the monetary motivation "to earn cash" but also by intrinsic motivations such as "to share beauty" and "to meet people." Using extensive transaction-level data, we find that hosts with intrinsic motivations post more property photos and write longer property descriptions, demonstrating greater engagement with the platform. Consequently, these hosts receive higher guest satisfaction ratings. Compared to hosts who want to earn cash, hosts motivated to meet people are more likely to keep hosting and to stay active on the platform, and hosts motivated to share beauty charge higher prices. As a result, these intrinsically motivated hosts have a higher customer lifetime value compared to those with a monetary motivation. We employ a multimethod approach including text mining, Bayesian latent attrition models, and lab experiments to derive these insights. Our research provides an easy-to-implement approach to uncovering consumer motivations in practice and highlights the consequential role of these motivations for firms.

Keywords: motivation, sharing economy, text mining, customer engagement, customer lifetime value, Airbnb

Jaeyeon (Jae) Chung (jc134@rice.edu) is an assistant professor of marketing at the Jones Graduate School of Business, Rice University, 6100 Main Street, Houston, TX 77005, USA. Gita Venkataramani Johar (gvj1@gsb.columbia.edu) is Meyer Feldberg Professor of Business in the Marketing Department at the Columbia Business School, 3022 Broadway, New York, NY 10027, USA. Yanyan Li (Yanyan.Li@marshall.usc.edu) is a PhD student in the Marketing Department at the Marshall Business School, University of Southern California, 3670 Trousdale Pkwy, Los Angeles, CA 90089, USA. Oded Netzer (onetzer@gsb.columbia.edu) is Arther J. Samberg Professor of Business in the Marketing Department at the Columbia Business School, 3022 Broadway, New York, NY 10027, USA. Matthew Pearson is a researcher, formerly at Airbnb (mpearson76@gmail.com). Please address correspondence to Jaeyeon (Jae) Chung. This article is based on essay 2 of the first author's dissertation at Columbia University chaired by the second author. The authors thank the review team and dissertation committee members Don Lehmann, Leonard Lee, and Silvia Bellezza. They thank the Chazen Institute of International Business for their support. Supplementary materials are included in the web appendix accompanying the online version of this article.

Editor: J. Jeffrey Inman

Associate Editor: Russell S. Winer

Advance Access publication 19 May 2021

 • Vol. 0 • 2021
DOI: 10.1093/jcr/ucab034

Motivation is fundamental to the study of consumer behavior. Scholars have made great strides in understanding the theoretical underpinnings and consequences of motivation and self-regulation. However, research has yet to make a strong case for the examination of consumer motivations by practitioners. There are several reasons for the gap between the emphasis on motivation research in the academic literature and its limited use in practice. First, it is not clear that studying consumer motivation is important for firms, because the effects of motivation have mostly been established in the short term. Usually, consequences are measured in the same time period as the manipulation or measurement of motivations (Herbst et al. 2012; Poynor and Haws 2009). The long-term effect of motivations in complex business contexts has been relatively understudied. Second, motivations have often been measured or manipulated in lab settings (Dommer, Swaminathan, and Ahluwalia 2013; Soman and Cheema 2004). Such measurements or manipulations often do not occur naturally and are difficult to replicate at scale and in field settings. Third, researchers often manipulate or measure one motivation at a time (Hong and Lee 2008) or examine motivational constructs at an abstract level, whereas multiple concrete motivations are likely to exist and operate together in practice. Given the difficulties in measuring the relationship between consumer motivations and firm-level outcomes, there has been little effort by firms to systematically measure and leverage consumer motivations at a granular level to improve long-term consumer–firm relationships.

The objective of this research is to extract consumer motivations from open-ended text responses and to examine the (possibly long term) downstream consequences of these motivations on actual consumer behavior in the context of the sharing economy. This context is particularly important because consumers (on both sides of the peer-to-peer platform) may participate in the sharing economy for different reasons and their motivations may affect how they behave on the platform. We study the relationship between consumer motivations and firm outcomes by utilizing a dataset obtained from Airbnb, one of the largest sharing economy platforms. On Airbnb, consumers work as both service providers (hosts) and service recipients (guests), and the firm makes a profit by connecting these two parties. We examine the relationship of hosts' motivation with host and guest behavior on the platform using transaction-level data obtained from Airbnb. We first use two text-mining machine learning approaches—a multilabel classification and a Poisson factorization topic model—to extract hosts' motivations to share their properties from responses of Airbnb hosts' to a single open-ended question regarding their motivations to host. After demonstrating convergent validity of extracting host motivations across these two approaches, we demonstrate how these motivations predict hosts' initial engagement as revealed by the text and photos of their home on the platform, guest satisfaction as measured by star ratings, and host long-term engagement with the platform as reflected by host activity, retention, pricing, and customer lifetime value. Two lab experiments provide evidence for the causal role of motivations in driving host behavior on the platform.

We find that Airbnb hosts are driven by several motivations that may operate singly or jointly. While the monetary motivation (earn cash) is commonly thought of as the main driver for participating in the sharing economy, we find that nearly 42% of the hosts are driven by nonmonetary motivations and nearly 64% of hosts with the earn cash motivation have at least one other motivation. Specifically, the most common nonmonetary motivations are to share the beauty of their homes and to meet people. Varying life circumstances such as frequent travel or empty nests may also motivate sharing. Across different analyses, we find that hosts who are motivated to share beauty and meet people, that is those who are intrinsically motivated, are likely to receive higher guest satisfaction ratings than cash-motivated hosts. This finding is robust even after controlling for a large set of property, host, and location characteristics. Furthermore, hosts who are motivated to share beauty charge the highest, and hosts motivated to meet people charge the lowest prices for their properties. Hosts who are motivated to earn cash price their properties in between these two groups. Overall, we find that hosts motivated to earn cash have the lowest Customer Lifetime Value (CLV).

Our research makes several contributions. First, we contribute to the growing literature on the sharing economy by shedding light on the potential nonmonetary motivations of hosts. Second, from a substantive point of view, this is one of the first papers to explore the relationship between motivations extracted from textual data and downstream consumer behaviors such as engagement and pricing decisions. Third, advertising from sharing economy companies often focuses on the economic benefits of participation. Our findings suggest that this singular focus on appealing to financial motivations may be suboptimal because the lifetime value of hosts with nonfinancial motivations is likely to be higher than that of hosts with monetary motivations. Fourth, from a methodological perspective, we extend past consumer behavior research on textual analysis (Humphreys and Wang 2018; Packard and Berger 2017) and propose the use of state-of-the-art machine learning and natural language processing approaches to extract motivations from textual data. These approaches overcome the limitations of numeric-scale-based methods and allow firms to identify consumer motivations at scale. This inductive method of identifying consumer motivations helps bridge quantitative and behavioral research in marketing (Berger et al. 2020).

In the rest of the article, we first discuss the limitations of commonly used paradigms to study the effects of motivation. Next, we discuss the emerging sharing economy, describe the Airbnb context, and present a conceptual model that examines the relationship between hosts' motivations and hosts' engagement with their guests as well as with the firm. In the empirical section of the article, we describe our dataset and use two text-mining approaches to extract and validate hosts' motivations from open-ended survey responses. After identifying the latent motivations, we run econometric models as well as experiments to test and quantify the relationship between hosts' motivation and guest satisfaction and hosts' long-term behavior. We conclude with a discussion of the implications of our findings.

MOTIVATION RESEARCH IN CONSUMER BEHAVIOR

Most motivation research in psychology and marketing follows an experimental paradigm that involves manipulating specific motivations such as promotion- and prevention-focus (Higgins 1998). Some commonly utilized motivation manipulations include a writing task (Dommer et al. 2013), article/scenario reading task (Cho and Johar 2011; Soman and Cheema 2004), or instructing participants to engage in a certain behavior such as entering a lottery to activate materialism-related motivation (Kim 2013). Researchers sometimes measure rather than manipulate individual differences in motivations using validated scales.

These manipulation and measurement approaches have limited use in field settings (Choy 2014; Dudwick et al. 2006). Many of the manipulations in experimental settings generate a momentary shift in respondents' motivations but cannot be routinely used by marketers and are often unnatural and/or difficult to implement in the field. Because the effects of manipulations are often short-lived (Queirós, Daniel, and Fernando 2017), the observed effects of motivations are bound to short-term perceptual and behavioral responses. Long-term behavioral consequences of motivations have been investigated only to a limited extent[1] because of the high monetary cost as well as the time-consuming nature of the procedure (Bauer 2004). In terms of numeric-scale measurement, firms often have difficulty asking a large population of consumers to respond to a set of long motivation scales (Choy 2014). Using such scales also requires a-priori knowledge of underlying consumer motivations (Bernard and Bernard 2013). In domains where new motivations emerge along with new modes of marketplaces (such as the sharing economy), validated numeric scales may not exist.

1 See Kivetz, Urminsky, and Zheng (2006) and Drèze and Nunes (2009) for notable exceptions.

To address the challenge of identifying consumer motivations that are unique to each firm, we propose that marketers collect and analyze consumers' textual responses (Roberts 2020). This information is commonly available due to the emergence of new communication channels such as review websites, personal pages, or blogs in which consumers freely express what they feel and think about products, services, and firms (Berger et al. 2020). Analyses of such large-scale textual data can uncover consumer motivations, especially in new economy settings such as the sharing economy where not all motivations have been fully uncovered. Collecting textual data is also relatively easy as consumers do not have to respond to long numeric-scale items and can decide what and how much to write. Additionally, textual responses can simultaneously capture multiple existing motivations in consumer minds, while manipulation and measurement often capture one motivation (or at best a few motivations) at a time.

To address the challenge of understanding the long-term downstream relationships between motivations and consumer behavior, we propose that firms incorporate their extracted motivations into traditional marketing analytics. By combining extracted motivations and transactional customer data, firms can better identify the impact of motivations and go beyond the common practice of considering motivations as primarily segmentation variables. In the next section, we describe our research approach in the context of the sharing economy in general, and Airbnb in particular.

MOTIVATIONS IN THE CONTEXT OF THE SHARING ECONOMY

Consumers who are motivated to share their possessions with others can now profitably do so by using platforms and marketplaces that facilitate sharing (Moorman et al. 2019). This so-called sharing economy is rapidly expanding from $15 billion in 2014 to a projected nearly $335 billion in 2025 (Tabcum 2019). While some sharing economy firms do not involve monetary transactions (e.g., Couchsurfing), most revolve around service platforms that allow consumers to become microentrepreneurs (Geissinger, Laurell, and Sandström 2020).

A good example of such a sharing economy platform is Airbnb, which is a home/room rental service that has more than 1.5 million rooms available in 220 countries (Airbnb 2020). Airbnb, similar to other sharing economy platforms, is based on the idea that consumers can become individual microentrepreneurs (Tabcum 2019; Zervas, Proserpio, and Byers 2017). Consumers on Airbnb can join the platform as hosts and/or as guests. Hosts can post the description of their property and add photos for potential guests to view before booking. After the trip is completed, both hosts and guests pay Airbnb for using its platform. A two-sided platform such as Airbnb needs to cater to both hosts and guests because they are both needed for the economic viability of the platform. As Airbnb writes on their website "Our platform is fueled by hosts and guests who are loyal to Airbnb because we treat them like members of a community, not commodities."[2]

Other notable examples of sharing economy platforms, where consumers are both service providers and service recipients include Uber, Lyft, Spinlister, Parkatmyhouse, and Relayrides. This business model is unlike the typical traditional economy model where consumers are defined as those who pay a fee to receive a product or service from the firm (Hamari, Sjöklint, and Ukkonen 2016). In the sharing economy, firms play an intermediary role and often receive service fees from both service providers and service recipients. Thus, both service providers and recipients are consumers of the firm (Viglia 2020).

This reliance on two types of consumers poses a challenge for the growing number of sharing economy platforms (Aspara and Wittkowski 2019; Chen and Wang 2019; Zervas et al. 2017). Firms often struggle to find ways to identify and motivate individual consumers to become microentrepreneurs and to keep them engaged on the platform. Consumers report that they are uncertain of the financial benefit they will gain from the hosting experience and are concerned about whether they will have positive experiences when interacting with guests (Yi, Yuan, and Yoo 2020). Given the relatively small service fee that firms receive from each transaction, scaling up operations is key, and firms rely heavily on the acquisition and retention of these microentrepreneurs. A critical question, therefore, concerns how to best motivate consumers to share their possessions and to remain engaged and active on the platform in the long run.

Motivations to Participate in the Sharing Economy

Consumers participate in the sharing economy for multiple reasons—to seek extra income, to resist and reject conspicuous consumption, to meet others, and to foster a collaborative community (Ozanne and Ballantine 2010). Other motivations include the better use of resource "slack" such as empty rooms, cars, and parking spaces and to redistribute goods (Benkler 2004; Lamberton 2015). Individuals on other sharing economy platforms, such as open-source software projects, also enjoy the feeling of competence and the development of collaborative knowledge outputs (Lakhani and Wolf 2003; Nov 2007). These intrinsic drivers are distinct features of the sharing economy that differentiate their providers from those in the traditional economy who provide goods and services primarily for financial benefits (Eckhardt et al. 2019; Lamberton 2016; van der Heijden 2004). For example, the European ride-sharing platform BlablaCar prides itself, as is evident from its name, with matching passengers and drivers for long-distance car rides; drivers and passengers share the objective of interesting companionship during the ride. However, most other sharing economy platforms such as Airbnb or Uber often induce people to join the platform by highlighting the monetary benefits of sharing their possessions. Our research questions whether this focus on monetary benefits is appropriate.

2 https://news.airbnb.com/airbnb-2019-business-update/

Motivations to Share Homes on Airbnb and Their Downstream Consequences

Besides economic benefits, which can be considered a form of extrinsic motivation, intrinsic motivations to enjoy the experience may also drive participation in collaborative communities (Hamari et al. 2016). According to self-determination theory (Deci and Ryan 1985; Ryan and Deci 2000, 2020), intrinsic motivations pertain to activities that are done to fulfill one's innate needs such as autonomy, competence, and relatedness. Extrinsic motivation, on the other hand, relates to activities done for reasons that have nothing to do with fulfilling one's needs or increasing one's satisfaction. Consumers who participate as Airbnb hosts may be driven by the extrinsic motivation to earn money or they may be intrinsically motivated, trying to fulfill innate needs such as relatedness, which concerns a sense of belonging or community. For example, hosts may enjoy meeting guests from different cultures and regions and derive value from cocreating local experiences with guests (Buhalis, Andreu, and Gnoth 2020; Hamari et al. 2016). We propose that host motivations—intrinsic or extrinsic—are likely to affect the way in which Airbnb hosts engage with their guests and with the firm.

Airbnb Host Motivation and Host Engagement. Drawing on self-determination theory, we propose a positive link between hosts' intrinsic motivation and their engagement with their guests. Relative to people who are extrinsically motivated, people who are intrinsically motivated are more immersed in their work, more committed, and more cognitively and emotionally invested in their tasks (Gruman and Saks 2011; Saks 2006). For example, fundraisers who had a stronger intrinsic motivation were more likely to make phone calls in a given week and raised more donations compared to those who were less intrinsically motivated (Grant 2008). These relationships between intrinsic motivation and performance have been demonstrated in various other domains including education and, most closely related to our application, hospitality (Grant 2008; Ryan and Deci 2000).

In the context of Airbnb, hosts' performance is measured by their ability to draw bookings as well as high satisfaction ratings from guests. To perform effectively, hosts have to begin by creating attractive posts of their property that include elaborate descriptions and photos. The amount of text and the number of accompanying photos on the platform reflect the host's devotion to the hosting task and can therefore serve as a measure of hosts' level of initial engagement on the platform (Karatepe and Karadas 2015). Consistent with the literature, we posit that intrinsically motivated hosts are likely to reveal greater initial engagement than extrinsically motivated hosts. Note that this form of initial engagement is meant to attract guests to book the property and it is likely to go along with other forms of host engagement such as the speed of response to inquiries, connecting with guests, and staying engaged with guests during their stay. These actions reflect one's commitment to providing a high quality of service (Harter, Schmidt, and Hayes 2002; Kumar, Lahiri, and Dogan 2018).

Airbnb Host Engagement, Guest Satisfaction, and Host Profitability. More engaged Airbnb hosts are likely to have more positive interactions with their guests, resulting in greater guest satisfaction. A positive relationship between service providers motivation and customer satisfaction has been found in the context of restaurants (Maylett and Warner 2014) and hospitality (Gundersen, Heide, and Olsson 1996; Jayaweera 2015; Spinelli and Canavos 2000). Highly engaged hotel employees have been found to exert effort to satisfy the needs of their guests and go beyond their duties to satisfy them (Karatepe 2013). Overall, research in the tourism and hospitality industry suggests that service providers' engagement level is a key factor in driving customer satisfaction (Borucki and Burke 1999; Tsaur, Hsu, and Lin 2019). Hosts with satisfied guests who leave flattering reviews are likely to draw other guests to their properties, have more bookings than those with less satisfied guests, and hence become more profitable hosts.

Host Engagement, Retention, and Profitability. Because intrinsically motivated hosts are more engaged and reap greater profits from hosting, they are also likely to remain active on the platform by opening up their homes for more bookings. They are also likely to remain on the platform in the long term because of the relative ease with which they generate earnings. Given the business model, whereby hosts pay a share of their revenues to the platform, the increased usage and retention of engaged hosts have a direct impact on platform profitability. These engaged hosts may also have an indirect impact on platform profitability because their guests become repeat customers of the platform (So, Kim, and Oh 2020).

Taken together, the greater margins derived from intrinsically motivated hosts, as well as their retention on the platform, are likely to translate to greater host CLV and higher firm profitability. CLV is defined as the net present value of all future streams of profits that a customer generates over the life of their business with the firm (Gupta, Lehmann, and Stuart 2004) and is a function of the customer's number of transactions with the firm, margins per transaction, and retention likelihood (Gupta and Lehmann 2006). CLV is a critical metric that firms can use to better manage their customer base and ensure that they are acquiring the right customers. We aim to show that uncovering host motivations can help in these efforts. While we have argued above that intrinsically motivated hosts are likely to be more engaged with hosting and are likely to stay with the firm longer (i.e., will show greater retention), the expected pricing of the property and margins per transaction are likely to depend on the specific type of intrinsic motivation as we demonstrate empirically.

In sum, we propose that Airbnb hosts who have intrinsic (vs. extrinsic) motivations exhibit greater engagement on the platform and garner higher satisfaction ratings from their customers. In turn, hosts with intrinsic (vs. extrinsic) motivations are more valuable to the firm because of their greater engagement with the platform as evidenced by their higher active levels and longer retention on the platform (figure 1). Additionally, more positive guest satisfaction in the short term contributes to greater host profitability in the long term, thereby creating a positive link between satisfaction and firm profit.

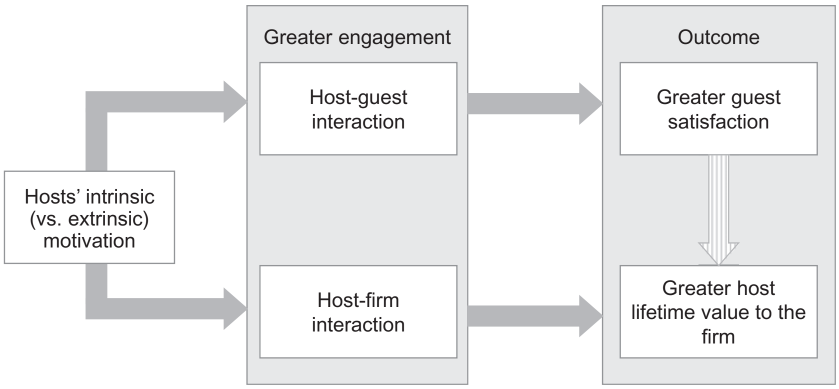
To test the model, we first extract host motivations from an open-ended text question on the Airbnb website. Then we examine whether intrinsically driven motivations correspond to greater host engagement and guest satisfaction than extrinsically driven motivations. Next, we test if hosts with intrinsic motivations are more active and stay engaged with the platform for longer than other hosts. Finally, we compare the CLV of intrinsically motivated hosts with that of extrinsically motivated hosts. In the next section, we describe the data that we use to explore Airbnb hosts' motivations.

DATA

Airbnb (www.airbnb.com) is one of the leading online sharing economy platforms where hosts share their properties with other consumers by listing their properties using text descriptions and photos. They can also specify the types of amenities (e.g., air conditioner, kitchen), the size of the properties (e.g., the entire home, private room), the price per night, and any additional surcharges (e.g., cleaning fee). Hosts can make their properties available for booking for certain nights while blocking other nights. After guests complete a stay at a host's property, they can provide a review of their experience by rating their satisfaction with the host on a five-star rating scale. Airbnb charges a guest fee of up to 14.20% of the total price guests pay for a stay and a host fee, which is about 3% of a host's earning (Airbnb 2021). This business model makes it important for Airbnb to attract both hosts and guests—hosts

FIGURE 1

CONCEPTUAL MODEL



Notes: Host–guest interaction refers to hosts' engagement with their guests as revealed by their property descriptions and host–firm interaction refers to hosts' engagement with the firm as revealed by their level of activity and retention on the platform. Guest satisfaction affects host lifetime value by increasing hosts' engagement with the firm.

need to provide satisfactory service in order for guests to use Airbnb and to report on their experience to attract other guests.

We use two primary datasets in this article. The first dataset comprises of the textual responses of 43,343 hosts who responded to an open-ended question ("Why did you start hosting?") posted on the Airbnb website from March 2013 to October 2014. The question was built into the site and popped up on the page when hosts logged into the site.[3] Note that this dataset is a one-time survey that captures host motivations, and we, therefore, treat host motivation as static.[4] Of the total 43,343 host responses globally, our text-mining analyses to uncover motivations focus on the subset of 22,842 hosts who have only one property listed on the platform. We removed hosts with more than one property to avoid professional hosts, such as property managers, whose motivations are likely to be different from those of lay individual hosts (Li, Moreno, and Zhang 2016). For the analyses using motivations as a predictor, we removed hosts outside of the USA and Canada so that we can use zip code level income and property value data to control for potential confounds as explained below. This left us with 13,337 hosts. We report the analyses of the global dataset of lay hosts (22,842 hosts) in web appendix B and show that the results are largely convergent.

The second dataset includes data on 291,746 reservation transactions (January 2011 to May 2015) of these 13,337 lay hosts (table 1). The dataset includes the following details of each completed reservation: (1) information about the reservation (e.g., the reservation date, the number of reserved nights, the total number of guests, price), (2) guests' review ratings, and (3) rich details of the host and property, such as location (e.g., town, country), the property type (entire home/private room), and a comprehensive list of amenities (e.g., air conditioner, doorman).

It is possible that motivations are confounded with the host's income as well as with the property value. For example, a finding that extrinsic motivations such as earning cash are associated with lower guest satisfaction could be attributed to characteristics of the host's property or host's socioeconomic status. We therefore control for these potential confounds by supplementing this dataset with additional data on zip code level property prices and household income. For each property, we know the city or town it is located in. We match that location with the corresponding five-digit or three-digit zip code for the U.S. listings, and the province or territory and census metropolitan area information for the Canada listings. For U.S. listings, we collected the Annual House Price Indices per zip code from the Federal Housing Finance Agency database (Bogin, Doerner, and Larson 2019) and the median household income per zip code from the U.S. Census database

3 The survey completion rate was approximately 30%. We compare our dataset with a control dataset of hosts who did not respond to the survey. We find that there are no significant differences between our sample and the control sample with respect to our key variables of interest, such as host engagement (e.g., listing number of words, listing number of photos, star ratings). We observe some differences in listing room types and the amenity count. See web appendix A for details.

4 We acknowledge the possibility that consumer motivations can be dynamic in terms of both strength and/or types. Analyzing changes in motivation over time is beyond the scope of the present research because of lack of data.

TABLE 1

HOSTS' TRANSACTIONAL DATASET: DESCRIPTIVE STATISTICS

N = 13,337 U.S. and Canada lay (non-professional) hosts	Mean	SD	Min.	Max.
Number of words used to describe a property	352.26	259.09	0	5,000
Number of photos used to describe a property	15.36	10.55	0	122
Listing max person capacity	3.34	2.11	1	16
Listing no. of bedrooms	1.39	0.89	0	10
Listing no. of amenities	12.23	4.11	1	30
The proportion of five-star reviews	0.75	0.25	0	1
Average reservation nights*	6.79	8.68	1.23	221.40
Average reservation guests*	2.77	1.66	1.23	19.68
Average Airbnb price within a city*	170.36	66.79	20.91	534.24
Zip code property price index	156.97	37.37	73.15	235.83
Zip code household income (in thousands)	77.33	17.20	26.83	152.38
The proportion of five-star reviews	75%	25%	0%	100%
Region				**Proportion**
USA				90.60%
Canada				9.40%
Room type				
Entire home/apt				61.06%
Private room				37.36%
hared room				1.58%

Notes: The top panel presents the descriptive statistics for hosts, listing properties, and reservations of the 13,337 lay U.S. and Canada hosts. We multiplied transaction-related variables (labeled with asterisk *) by a factor that is randomly drawn from Uniform (0.5, 1.5) to disguise the actual values for confidentiality reasons.

(Census 2021). We included the zip code level average property value and income over the years 2011–2015 for each U.S property. Similarly, for the Canadian listings, we collected the New Housing Price Index and median household income from the database of Statistic Canada (Government of Canada, Statistics Canada 2021). We mapped this information onto the property listings based on the province or census metropolitan area information and averaged over years (2011–2015). As discussed later, we also controlled for local market price on Airbnb.

In the following section, we leverage text mining and machine learning tools to extract host motivations from open-ended text and use econometric and statistical analyses as well as lab experiments to investigate the relationship between host motivations, engagement, guest satisfaction, and firm profitability. Figure 2 outlines our multimethod approach.

UNCOVERING MOTIVATIONS

Two Approaches to Extract "Motivations" from Text

There are two approaches one can take to extract (latent) motivations from text. First, when researchers have prior knowledge or assumptions about possible consumer motivations and are willing to incorporate this knowledge to guide their machine learning exploration, *supervised* machine learning methods are appropriate. For example, a firm may provide inputs based on internal analysis, interviews, or focus groups, which can be used to train supervised learning models. By contrast, when researchers want to automatically explore latent (and often unknown to the researcher) traits from textual data, *unsupervised* machine learning methods are appropriate. For example, in new or emerging domains, or when researchers have no access to the firm's internal knowledge, this method can be particularly useful.

In our application, we employed and compared both approaches. We used insights obtained from Airbnb on what they believed were the possible set of motivations and trained a supervised machine learning model—multilabel classification—to classify hosts' motivations. Additionally, we utilized an unsupervised topic modeling—Poisson factorization approach, to uncover latent hosts' motivations. We then assessed convergent validity between these two methods. Prior to running the two machine learning models, we preprocessed the textual data following standard steps in text analysis. The details of the text preprocessing are described in web appendix C.[5]

Supervised Machine Learning: Multilabel Classification

Airbnb provided us with some information on the potential set of motivations identified by their consumer insights team (table 2). Leveraging this prior knowledge, we applied a multilabel classification to identify existing motivations at the host level. Because a host can have more than one motivation (e.g., a host can be motivated to both earn cash and to meet people), we use the multilabel classifier, which allows for multiple motivations simultaneously, as opposed to binary classifier.

Method. The first step in building a classifier is creating a training dataset. We need to label a subset of the textual data on the predefined motivations and then use these to predict classification of motivations in the entire dataset. To label the training dataset, we recruited 170 Mechanical Turk human coders (MTurkers) to read and classify a

5 To increase the size of the training data for the textual analysis, we use the text responses from all hosts (globally) who had only one property N = 22,842.

FIGURE 2

FLOW OF ANALYSES

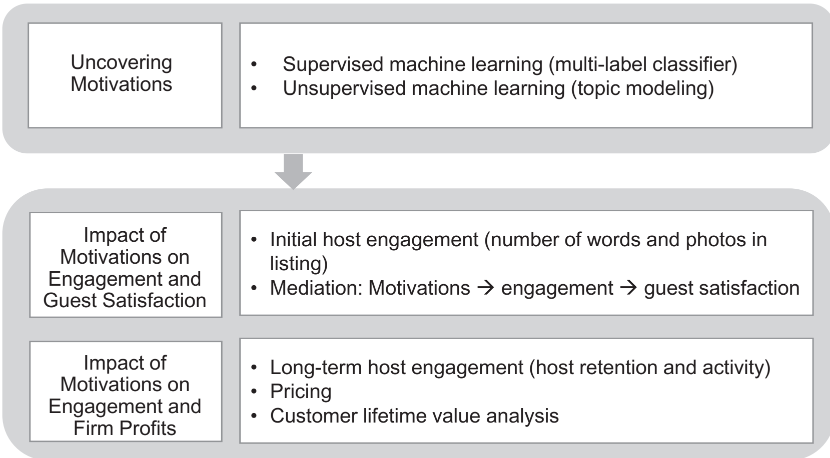


TABLE 2

PREDEFINED MOTIVATIONS BY AIRBNB

Motivation	Definition
Earn cash	People host to aid their financial needs.
Meet people	People enjoy meeting guests from all over the world and from different cultures. They like to make friends.
Share beauty	Sharing the beauty of a nicely maintained home, a beautiful city, and surroundings gives happiness. People derive pleasure from giving recommendations and serving the needs of their guests.
Life circumstances	People encounter life circumstances (ex: they travel a lot, have an empty room/space, a roommate moved out, etc.), and use this as an opportunity to host.
Pay forward	People had a great experience as a guest and now want to also offer a similar experience by becoming a host
Recommendation from others	Friends, family, and other people who have been hosting recommended them to join and serve as a host.
Other motivations	Other reasons that are not identified above.

sample of hosts' text responses using the seven motivations predefined by the company (table 2). Human coders are often used as a reliable source to interpret the meaning of text (Liu et al. 2012; Ou, Khoo, and Goh 2008).

Each MTurker was paid \$0.70 to evaluate the textual responses of 20 hosts who were randomly sampled from a sample of 527 hosts,[6] and this was used as the training dataset. MTurkers were provided with information about Airbnb and the predefined motivations. They were asked to read each of the host's text responses and categorize whether any of the seven motivations was mentioned in each text. Note that each text can be labeled with more than one motivation. We randomly assigned hosts' text to MTurkers such that each host's text would be evaluated by a minimum of three MTurkers. Each host's text was coded as 1 for a specific motivation (otherwise 0) if 50% or more of the MTurkers indicated that the motivation was mentioned in the text (inter-rater coder agreement: $\alpha = .79$; see web appendix D for details of the classification survey).

Next, we used the human-labeled dataset to train the multilabel classifier. We used the words[7] in each host's textual response to predict whether each of the seven

6 The original training dataset included a total of 1,000 host responses, but after removing hosts with more than one property (professional hosts), we were left with 527 hosts.

7 The occurrence of a word is based on the term frequency-inverse document frequency (Tf-Idf) measure (Salton and Buckley 1988). Tf-Idf is a commonly used measure in information extraction, which measures the weighted occurrence of a word in a document given the frequency of the word appearing in all documents in the corpora and the length of the document.

motivations was present in the text. We applied the Binary Relevance method (Tsoumakas and Katakis 2007), which decomposes a multilabel learning task into multiple independent binary classification tasks, fitting one binary classifier for each motivation against all the other motivations, and used Support Vector Machine with linear kernel to train the multilabel classifier.[8] To avoid overfitting and to compare with other classifiers, we ran a five-fold cross-validation to calibrate the classifier. After the classifier was calibrated on the set of 527 responses, we applied it to the entire set of textual responses of lay hosts. This enabled us to uncover whether each host had one or more of the seven motivations.

Results and Discussion. Table 3 summarizes the performance of the supervised multilabel classification model. The second column shows the proportion of each motivation in the human-labeled dataset, which can serve as a proxy for the "true" proportion of each motivation across hosts. Note that the proportions do not add up to 100% because each host can have more than one motivation (see web appendix E for the overlap across motivations and the correlation between these motivations). We find a high degree of agreement between the aggregate share of motivations in the human-coded data and those based on the multilabel classifier (columns 2 and 3 in table 3). The three motivations, "earn cash," "share beauty," and "meet people," have the highest levels of predictive accuracy with area under the curve (AUC) above 0.84. The other motivations also have AUC greater than 0.7, suggesting a statistically acceptable prediction accuracy. The in-sample and out-of-sample hit-rate, averaged across the five-folds, also show high reliability of the classifier. Similarly, strong predictive accuracy is revealed using the Jaccard index (e.g., Netzer et al. 2012; Toubia and Netzer 2017), which focuses on the ability of the algorithm to accurately predict the existence (as opposed to the absence) of a motivation.

Overall, the supervised learning approach to explore host motivations shows a high degree of accuracy. One of the limitations of this approach, however, is that it requires prior knowledge and a relatively large sample of human-classified text. The next section presents an unsupervised machine learning method that does not require prior knowledge (from the firm) to guide the learning process. Using this approach can help validate the firm's prior knowledge and aids in possibly uncovering additional motivations without a need for human labor in classifying the text.

8 The classification is conducted using the scikit-learn 0.18 package in Python (http://scikit-learn.org/).

Unsupervised Machine Learning: Topic Modeling Using Poisson Factorization

We adopted a topic modeling approach that examines the combinations of words to infer latent topics (e.g., motivations). We chose the Poisson factorization topic modeling approach because it has several advantages in our application relative to other approaches (e.g., Latent Dirichlet Allocation; LDA). First, Poisson factorization is well suited for sparse and short textual responses (Canny 2004). Second, unlike LDA, Poisson factorization does not assume that the distribution of topics in a document sums up to 1. That is, Poisson factorization allows some documents (text responses) to include more topics (motivations) than others, which is a desirable feature in our application.

Method. Poisson factorization assumes that each text is constructed of a mixture of multiple topics with a Gamma prior and that the occurrence of words in each text is probabilistically related to each topic with a Poisson distribution (Canny 2004). We used the variational inference algorithm to estimate the model (Gopalan, Hofman, and Blei 2013). See web appendix F.1 and F.2 for details of the Poisson factorization model and the variational inference estimation.

Results and Discussion. The first step in topic modeling is selecting the number of topics and evaluating the quality/interpretability of these topics. Considering both model fit—the perplexity score (Blei, Ng, and Jordan 2003)—and interpretability of the topics/motivations, we chose seven topics (see web appendix F.3 for details). These topics are (the proportion of appearance of each topic across hosts is in parentheses): earn cash (39%), meet people (33%), share beauty (24%), life circumstances (33%), recommendations from others (29%), resource utilization (30%), and other motivations (25%).[9] Each topic can be interpreted by the most common words associated with that topic (table 4). The top words associated with each motivation are presented based on the relevance measure (Sievert and Shirley 2014), which balances the frequency and uniqueness of each word in each topic.

Based on the words in each textual response, Poisson factorization assigns a score for the occurrence of each motivation in each textual response. For example, for the response "I want to earn money and make friends from around the world," the Poisson factorization analysis would generate the scores of each motivation (i.e., Poisson posterior means) in survey responses (e.g., earn cash: 1.95,

9 "Resource utilization" and "Life circumstances" seem similar, however, examining the words associated with these motivations reveal different semantic meaning. "Resource utilization" relates more to time and space resources, whereas "Life circumstances" relates more to making money due to a life opportunity. From a statistical point of view, the correlation between the "Resource utilization" and "Life circumstances" is quite low ($r = .105$).

TABLE 3

PERFORMANCE OF THE MULTILABEL CLASSIFICATION IN EXTRACTING HOST MOTIVATIONS.

Motivation	Human label percentage ($N = 527$) (%)	Multilabel classification percentage ($N = 22{,}842$) (%)	AUC	Correlation with human coding	Hit-rate (%)		Jaccard index	
					In-sample	Out-of-sample	In-sample	Out-of-sample
Earn cash	58	51	0.94	0.97	99	87	0.97	0.78
Share beauty	20	22	0.84	0.93	98	83	0.90	0.41
Meet people	37	35	0.92	0.97	99	85	0.98	0.66
Life circumstance	42	44	0.81	0.92	96	73	0.91	0.52
Paying forward	16	17	0.81	0.92	98	80	0.92	0.30
Recommendations from others	7	4	0.78	0.99	99	94	0.99	0.13
Other motivations	13	10	0.66	0.95	99	83	0.93	0.10

Notes: The second column shows the proportion of hosts who have a specific motivation in the human-labeled dataset. For example, out of the 527 hosts in the human-labeled dataset, 58% have the "earn cash" motivation, and 20% have the "share beauty" motivation based on human coding. The third column shows the proportion of hosts out of all the 22,842 host responses that are predicted to have each motivation. AUC of receiver operating characteristic curve, a measure commonly used for prediction accuracy of binary outcomes is used for the hold-out prediction accuracy. The hit-rate is the percentage of motivations that are correctly predicted by the classifier. The Jaccard index is defined as ("number of correctly predicted motivations")/("the number of motivations that are labeled by human coders but missed by the prediction from the classifier" + "the motivations that are predicted by the classifier but are not labeled by human coders and are correctly predicted motivations").

TABLE 4

POISSON FACTORIZATION: THE TOPICS AND THEIR ASSOCIATED TOP WORDS

Motivations (topics)	Top 10 words by relevance (relevance factor = 0.5)
Earn cash	Money, extra, make, cash, earn, little, income, spare, meet, room
Meet people	People, love, new, meeting, like, world, meet, enjoy, around, different
Share beauty	Share, beautiful, others, lovely, large, cottage, home, area, unique, beach
Life circumstances	Apartment, time, empty, rent, away, lot, long, term, flat, travel
Recommendations from others	Friend, try, guest, first, site, decided, service, friends, easy, heard
Resource utilization	Help, pay, vacation, cover, costs, income, holiday, expenses, mortgage, bills
Other motivations	Stay, staying, hotel, place, city, offer, local, provide, hotels, accommodation

meet people: 1.55, share beauty: 0.12, resource utilization: 0.15, life circumstances: 0.10, recommendations from others: 0.07). To compare the multilabel classifier approach to the probabilistic unsupervised approach, we need to assign motivations to hosts. We assign the motivation to the host if the Poisson posterior mean for the motivation for the host is greater than the average across motivations and hosts (note that following this approach each host can have more than one motivation).[10] The next section examines the convergence between the supervised and unsupervised machine learning models.

Convergence of the Supervised and Unsupervised Machine Learning Approaches

Five out of the six motivations suggested by the firm (not including "other motivations") overlapped with our unsupervised approach. The topic "paying forward" (which was suggested by the firm) was fairly small in our human coding, and it was not recovered in the unsupervised approach. Instead, Poisson factorization uncovered another motivation—resource utilization.

We used several approaches to assess the convergence of the two motivation extraction approaches. First, we examined the correlation of the Poisson factorization topics with multilabel classification and the human-labeled data based on the company motivation labels (see columns 3 and 4 in table 5). We see a fairly high correlation between the two approaches. Additionally, we assessed AUC, Hit-rate, and Jaccard Index (columns 5–7 in table 5) between the Poisson factorization and the human-labeled data. We observe AUC and hit rates both at the 70% range or above, again suggesting strong convergence between the Poisson factorization approach and human coding of the motivations. Specifically, for three motivations ("earn cash," "share beauty," "meet people"), the two machine learning approaches produce a particularly high degree of

10 Using the continuous motivation scores instead of assigning motivations to hosts results in a consistent pattern of results.

TABLE 5

CONVERGENCE OF POISSON FACTORIZATION

Motivation	Topic Consensus (Human evaluation hit-rate)	Correlation with multilabel classification	Compare with human-labeled data					
			CorrelationAUC			Jaccard index	Average of posterior probability	
					Hit-rate (%)		Human label= 1	Human label= 0
Earn cash	0.87	0.54	0.53	0.7674		0.55	0.59	0.07
Life circumstances	0.60	0.47	0.42	0.7072		0.47	0.58	0.18
Meet people	0.80	0.58	0.57	0.7880		0.57	0.70	0.14
Share beauty	0.83	0.45	0.51	0.7782		0.44	0.66	0.13
Recommendations from others	0.74	0.24	0.26	0.7475		0.16	0.74	0.25
Resource utilization	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Notes: The second column shows the percentage of MTurkers who correctly identified the predefined labels after reading the top 20 words from each topic in Poisson factorization. The third column shows the correlation between the motivations extracted from Poisson factorization and that extracted from the multilabel classification. From the fourth column and on, we use the MTurk human-labeled dataset to compute the correlation, AUC, hit-rate, Jaccard index, and the average of posterior probabilities of motivations extracted from Poisson factorization (also see "note" in table 3 for details). Resource utilization is marked as "N/A" because it is a motivation identified by Poisson factorization but not identified by the Airbnb classification. The two rightmost columns show the probability of each topic based on the Poisson factorization split by whether the topic was identified in the text by human coders or not..

convergent validity. Additionally, the "topic consensus" metric commonly used in the topic modeling literature (Morstatter and Huan 2017) suggested that humans (MTurkers) were able to consistently match the words with the highest relevance for each topic in the unsupervised learning approach (table 4) with the label provided by the firm (table 2). As seen in the second column of table 5, the percentage of MTurkers who correctly identified the predefined labels after reading the top 20 words from each topic in Poisson factorization is generally high. See web appendix F.4 for details of the topic consensus analysis.

Overall, the three topics "earn cash," "meet people," and "share beauty" have high convergence validity across multiple convergence metrics. The "life circumstances" motivation has somewhat lower convergent validity but it frequently appeared in the text (see columns 2 in table 3). Other motivations (recommendation from others, resource utilization) appeared infrequently in the text. Thus, in our subsequent analyses, we focus on the four main motivations uncovered by the two approaches, "earn cash," "share beauty," "meet people," and "life circumstances." Note that "earn cash" and "life circumstances" relate to extrinsic motivations whereas "share beauty" and "meet people" relate to more intrinsic motivations. In addition, we also explore the most commonly appearing pairs of motivations to see if there are any interaction effects of these motivations ("earn cash × meet people," "earn cash × life circumstances," "meet people × share beauty"). The next section examines the downstream consequences of different motivations extracted from the theory-based multilabel classifier.[11]

IMPACT OF MOTIVATIONS ON ENGAGEMENT AND CUSTOMER SATISFACTION

Airbnb Data: Motivation and Initial Host Engagement

We measure hosts' initial engagement (i.e., at the time of joining the Airbnb platform) using the number of words and photos in their property listings. These are appropriate variables to measure engagement because hosts advertise their property through text and images (Zhang et al. 2017). This is analogous to the idea that verbally introducing and guiding guests of a hotel facility during their stay represents an employee's level of engagement (Blue and Harun 2003). We predict that intrinsically motivated hosts—those motivated by "sharing beauty" and "meeting people"—are likely to be more engaged than those who are extrinsically motivated to "earn cash" (Deci, Koestner, and Ryan 2001; Deci and Ryan 1985).

Table 6 reports model-free evidence for the relationship between motivations and initial engagement and reveals that hosts who are motivated to share beauty or to meet people are more engaged than hosts who do not have these motivations. Hosts who are motivated to earn cash are significantly less engaged than hosts who do not have that

11 As a robustness check, web appendix G repeats the analyses using the motivations extracted using the Poisson factorization. Across all these analyses, the relationship between motivations and downstream consequences is highly consistent with results reported in the main text.

motivation. For hosts with the life circumstances motivation, the results are mixed.

To test for the significance of this effect while controlling for other variables, we regressed the engagement variables on the four motivations (and the most common interactions between motivations[12]) while controlling for a large set of property, host, and location characteristics that may differ by motivation (see table 1 for the summary of key variables and web appendix H for the full list of property characteristics by motivation). Specifically, we controlled for room types (i.e., entire home, private room, shared room), whether the host is verified and whether the host has a profile photo. We also aggregated reservations at the host level and controlled for the average number of nights the property was booked, guests per reservation, and the average reservation price per night. As stated earlier, we controlled for the property price index and median household income (in thousands) in the property zip code. For local market price, we controlled for the average listing price of the same type of property (i.e., entire home/private room/shared room) on Airbnb (from our Airbnb dataset) in the city or town that the property was located in, as well as city fixed effects for all cities that had at least 10 listings ($N = 247$ cities). The smaller regions that had fewer than 10 listings were labeled as "other" and were used as a baseline of the fixed effect.

Lastly, we controlled for 33 property amenities (e.g., TV, internet, AC, kitchen, gym, breakfast, heating, washing, pets, fireplace). Because the property amenities are often highly correlated, we reduced the dimensionality of the amenities using LDA analysis to three main amenity topics—"basic amenities," "safety-related amenities," and "apartment complex amenities." See details of the topics in web appendix I.

Consistent with the model-free results in table 6, we find that hosts who are motivated to share beauty and to meet people write significantly more words and post more photos in describing their property than hosts who do not have these motivations (table 7). This effect is robust after controlling for a strong set of property, host, and location characteristics including city fixed effects. We did not uncover similar positive effects for hosts who had the earn cash or life circumstances motivations. We also did not find significant interaction effects for the three pairs of motivations. The lack of significant interactions between the motivations suggests that one can capture the effect of multiple motivations by simply accounting for the effect of the presence of each of the motivations.

12 We included the interactions earn cash and meet people; earn cash and life circumstances; and meet people and share beauty. We also repeated our regression using the combination of $2^4 = 16$ interaction pairs in the above analyses and did not find statistically significant effects for the other (less frequent) interaction terms on the engagement variables.

In sum, consistent with our conceptualization, the two intrinsic motivations—share beauty and meet people—are significantly associated with higher levels of host engagement as reflected in the number of words and photos in their listings. However, despite our attempt to control for a large set of observed host and property characteristics, motivations may still be correlated with unobserved property or host characteristics. We address this concern by conducting a lab experiment to test the causal relationship between intrinsic versus extrinsic motivations and initial engagement. Due to the ambiguity in the meaning of different "life circumstances," the experiment only manipulates the three motivations of key interest (earn cash, meet people, share beauty).

Experiment 1: Motivation and Initial Host Engagement

This experiment aims to demonstrate the relationship between motivations and engagement by manipulating host motivation and measuring host engagement, holding property characteristics constant. Consistent with the results of our secondary data analyses, we expect that intrinsically motivated participants (those hosting in order to meet others or to share beauty) are likely to use more words to describe their homes compared to those who are extrinsically motivated (those hosting to earn cash).

Method. Prolific participants from the U.S. ($N = 165$; $M_{age} = 32.32$, 40.00% male) first read a brief introduction to the company, Airbnb, as an online platform that allows people to advertise rooms in their homes and receive bookings. The cover story asked participants to imagine living in Chicago and renting out a room in their home on Airbnb. The scenario continued by asking them to imagine that they were considering renting out the room for one of the following reasons—to earn cash, to meet people, or to share beauty. In order to elicit one of these motivations, we created a 70-word description of each motivation (see web appendix J.1 for details) based on the words extracted from the Poisson factorization (table 4). This manipulation was pretested and it successfully elicited one of the three motivations (see web appendix J.2 for the pretest).

All participants then saw the same set of property images taken from a real property on Airbnb (see web appendix J.3 for the property image) and were told that their goal was to rent the room shown in the picture by posting a property advertisement on Airbnb. All participants were then asked to write an advertisement as if they were listing their room on Airbnb. Participants then reported their level of effort in the writing task and answered some demographics questions at the end.

Results and Discussion. Based on the TaskMaster Qualtrics toolkit that captures distraction by measuring the amount of time spent outside of the survey website

TABLE 6

HOST ENGAGEMENT AS REFLECTED IN PROPERTY DESCRIPTIONS

	Earn cash		Meet people		Share beauty		Life circumstances	
	Mean	STDEV	Mean	STDEV	Mean	STDEV	Mean	STDEV
Number of words	343.25	245.52	384.59	272.39	388.57	295.25	359.47	256.17
Number of photos	14.93	9.94	15.78	10.89	16.58	11.64	15.48	10.64
	t-stats	*p*-value	*t*-stats	*p*-value	*t*-stats	*p*-value	*t*-stats	*p*-value
Number of words	−4.406	.000	10.256	.000	8.980	.000	2.989	.003
Number of photos	−5.195	.000	3.281	.001	7.371	.000	1.170	.242

Notes: The means are from all hosts who have each motivation. These hosts may or may not have other motivations. The results are consistent when we only examine hosts with a single motivation. *t*-stats compare hosts who have a given motivation (e.g., hosts who have the earn-cash motivation) with those who do not have the motivation (e.g., hosts who do not have the earn-cash motivation).

(Permut, Fisher, and Oppenheimer 2019), two participants who spent more than five minutes away from our survey, and four participants who were not in the age range of 18–60 years old, were excluded from the analyses. First, we conducted a text analysis of the advertisements created by participants using the LIWC dictionary (Pennebaker et al. 2015) to examine if the manipulation successfully elicited each of the three motivations. Results confirmed that respondents in the earn cash condition used more money-related words in their description, respondents in the meeting people condition used more affiliation-related words and those in the share beauty condition used more feeling-related words (see web appendix J.4 for details).

An ANOVA revealed a significant main effect of motivation ($F(2, 156) = 3.56$, $p = .031$). As expected, compared to those who were motivated to earn cash ($M_{\text{earn cash}} = 48.60$, SD = 23.75), participants who were motivated to meet people ($M_{\text{meet people}} = 66.13$, SD = 41.76), and those who were motivated to share beauty ($M_{\text{share beauty}} = 65.70$, SD = 42.44) used significantly more words to describe their property in the ad (earn cash vs. meet people; $F(1, 156) = 5.59$, $p = .019$, $d = .52$; earn cash vs. share beauty, $F(1, 156) = 5.36$, $p = .022$; $d = .50$). There were no significant differences between the meet people and share beauty conditions ($F(1, 156) = 0.004$, $p = .952$, $d = .01$).

A possible confound in our analysis is that participants in the two intrinsic motivation groups may have merely copied-and-pasted the words from the manipulation in the survey to create their hosting ad. The length of the manipulation was identical across motivation conditions, thus "cutting and pasting" should not reveal differences across conditions. However, it is possible that participants in the earn cash motivation group would be less likely to copy-and-paste the language from the manipulation because it does not lend itself to an ad. To test this alternative account, we calculated the proportion of words shared between the motivation manipulation and participants' ad descriptions. The proportion of shared words was generally low ($M_{\text{overall}} = 19.96\%$), suggesting that participants did not merely copy-and-paste the manipulation text. In fact, over 80% of words were unique. Despite the generally low means, we did find, however, that participants in the meet people condition ($M_{\text{meet people}} = 21.32\%$, SD = 0.12) and the share beauty condition ($M_{\text{share beauty}} = 23.41\%$, SD = 0.12) used significantly more words that also appeared on the manipulation, compared to those in the earn cash condition ($M_{\text{earn cash}} = 14.37\%$, SD = 0.08; respectively, $p < .001$, $p = .001$). We therefore calculated the cosine similarity score between the word occurrence in the participant's ad description and the motivation manipulation and controlled for it as a covariate ($p = .262$) in the analyses. The effect of motivation on the number of words in the ad remains significant ($F(2, 155) = 4.18$, $p = .017$) after controlling for this covariate.

We also examined the time-to-completion of writing the ad and found no statistically significant differences across the conditions, suggesting it is unlikely that the intrinsically motivated groups copied the stimuli and the earn cash group did not ($M_{\text{meet people}} = 488.20$ seconds vs. $M_{\text{share beauty}} = 453.05$ seconds vs. $M_{\text{earn cash}} = 452.75$ seconds; $p = .710$). Overall, these results provide evidence for the causal impact of motivations on engagement. The next section examines our proposition that intrinsic host motivations that are associated with higher engagement, are also associated with higher satisfaction as reflected in guest ratings.

Effect of Host Motivations on Guest Satisfaction Mediated by Engagement

Guests can provide satisfaction ratings on a scale of five stars after completing their stay at hosts' properties. These guests may not have been directly aware of the hosts' motivation at the time of rating their satisfaction. However, intrinsic motivation manifests in greater host engagement as observed in the level of detail in their property listings and may also be reflected in other online and offline host–guest interactions. Hence, we postulate that the two intrinsic

TABLE 7

RELATIONSHIP BETWEEN MOTIVATIONS AND HOST ENGAGEMENT, CONTROLLING FOR PROPERTY, AND HOST CHARACTERISTICS

	Number of words		Number of photos	
	B	SE	B	SE
Earn cash	−3.251	(7.301)	−0.134	(0.291)
Meet people	34.200***	(7.930)	1.047***	(0.316)
Share beauty	24.516***	(7.305)	0.879**	(0.291)
Life circumstances	9.789	(6.513)	−0.116	(0.259)
Earn cash × Meet people	3.288	(9.431)	0.005	(0.375)
Earn cash × Life circumstances	5.929	(8.770)	0.199	(0.349)
Meet people × Sharing beauty	−5.27	(10.582)	−0.418	(0.421)
Host is verified	49.504***	(4.604)	1.576***	(0.183)
Host has a profile picture	29.546***	(4.413)	0.953***	(0.176)
Hosts' total booking as guests	2.190***	(0.642)	0.091***	(0.026)
Listing person capacity	14.814***	(2.027)	0.809***	(0.081)
Entire room	26.226	(21.929)	3.971***	(0.873)
Private room	38.038*	(17.162)	2.810***	(0.683)
Cancellation policy	29.029***	(1.925)	1.097***	(0.077)
Allows for long rental	36.132***	(7.391)	1.999***	(0.294)
Instant booking available	86.113***	(8.116)	4.421***	(0.323)
Email contact prior to guest arrival	1.133**	(0.371)	0.007	(0.015)
Average reservation nights	−1.750***	(0.368)	−0.052***	(0.015)
Average reservation guests	−5.305	(3.247)	0.084	(0.129)
Safety-related amenities	314.152***	(38.510)	6.926***	(1.533)
Basic amenities	−72.436*	(36.201)	−3.290*	(1.441)
Average price within a city	−0.172	(0.139)	−0.009	(0.006)
Average price per night	−0.043	(0.025)	0.006***	(0.001)
Average price within a zip code	1.349*	(0.635)	0.002	(0.025)
Average income within a zip code	−0.961	(0.840)	0.033	(0.033)
Country effect: Canada vs. USA	81.519*	(40.578)	0.840	(1.615)
Constant	5.958	(95.811)	1.919	(3.813)
City fixed effect	Yes		Yes	

Notes: The apartment complex amenity is omitted and serves as a baseline for the two other amenities variables—basic amenities and safety-related amenities. "Average price within a city" is the average listing price of the same type of property (i.e., entire home/private room/shared room) on Airbnb in the city or town. "Average price per night" is the listing price for reservation on Airbnb. "Average price within a zip code" is the property price (i.e., the home value) within a zip code. We also checked the Variance Inflation Factor (VIF) to detect multicollinearity among the regressors. The VIF of all regressors from the transactional dataset is less than 2 in the above two regressions. Overall, these results suggest a relatively low level of multicollinearity among the regressors. $N = 13{,}337$ ***$p < .001$, **$p < .01$, and *$p < .05$.

motivations (meet people, share beauty) are likely to lead to higher guest satisfaction ratings. Note that our measures of initial host engagement (number of words and photos in the listing) serve as a proxy for host engagement in terms of their interaction with guests. We suggest that host motivations are likely to affect guest satisfaction as a result of engagement.

Effect of Host Motivations on Guest Satisfaction Mediated by Engagement. To test this idea, we ran a mediation analysis using each one of the motivation variables as a predictor, engagement variables (# of words/photos) as parallel mediators, and the proportion of five-star reviews as a dependent variable.[13] The same set of control variables[14] from the previous analyses were used as covariates.[15] First, we found significant and positive direct effects of both the meet people ($B_{\text{meet people}} = .022$; $p < .001$) and share beauty ($B_{\text{share beauty}} = .016$; $p < .001$) motivations on proportion of five-star reviews. The earn cash ($B_{\text{earn cash}} = .002$; $p = .673$), and the life circumstances motivations revealed insignificant relationships ($B_{\text{life circumstances}} = .006$; $p = .194$). Next, we examined if the positive effect of the meet people and share beauty motivations on the proportion of five-star ratings was mediated by hosts' engagement (parallel mediation PROCESS Model 4). We used a 95% bias-corrected confidence interval based on 5,000 bootstrap samples. The reported results are unstandardized coefficients as each variable represents meaningful scales (Pieters 2017). Results revealed that when controlling for the identical set of covariates, having (vs. not having) the motivation to meet people was positively mediated both by the number of words (Indirect effect = .0011, CI = [.0002, .0021]) and by the number of photos (Indirect effect = .0024, CI = [.0015, .0034]) in predicting the proportion of five-star guest ratings. We found a similar pattern of results for the share beauty motivation (the number of words: Indirect effect = .0009, CI = [.0002, .0017]; the number of photos: Indirect effect = .0018, CI = [.0010, .0028]). Finally, controlling for the two engagement variables, the effect on the proportion of five-star ratings remained significant for the meet people motivation ($B'_{\text{meet people}} = .019$; $p = .001$) and the share beauty motivation ($B'_{\text{share beauty}} = .014$, $p < .01$). For the visualization of these mediation models, see web appendix L.

As discussed previously, the host engagement variables—the number of words and photos in the property listing—indicate host engagement level when attracting guests through the property description, and not necessarily

13 We used the proportion of five-star reviews rather than average ratings because star ratings on Airbnb highly skewed to five stars. Such high levels of skewedness are often observed in online ratings and particularly in sharing economy platforms (Schoenmueller, Netzer, and Stahl 2020). See the distribution of star ratings in our dataset in web appendix K.1.

14 We excluded region fixed effects because it is infeasible to run the PROCESS model with such a large number of dummy variables. Instead, we still controlled for the average property price of the regional market.

15 Web appendix K.2 reports the regression of the proportion of five stars in satisfaction subratings (accuracy, cleanliness, communications, check-in, value, and amenities) on motivations, controlling for the same set of covariates used in the previous analyses.

the engagement level of hosts *during* each guest's visit. Yet the finding that these engagement variables drive guest satisfaction suggests that these variables also capture hosts' general levels of engagement with their guests.

The next section examines the possible long-term effect of host motivation on their engagement with the firm as measured by host activity and retention on the platform. We also examine the relationship between guest satisfaction and host profitability.

THE RELATIONSHIP BETWEEN MOTIVATIONS AND PROFITABILITY

The host's lifetime value depends on the margins they generate that are based on their activity on the platform (number of bookings) as well as the price charged. The host's lifetime value is also influenced by how long hosts are active on the platform that is their retention by Airbnb. As discussed earlier, we posit that intrinsically motivated hosts are likely to be more active on the platform and stay longer, leading to a higher CLV than extrinsically motivated hosts. Based on the specific motivations uncovered for Airbnb, we also propose that intrinsically motivated hosts who are motivated by "sharing beauty" are likely to price their properties higher than those motivated to "meet people." This is because homeowners who put a subjective and emotional valuation on top of the objective valuation (e.g., those who want to share beauty) are likely to inflate the price (Genesove and Mayer 2001). On the other hand, hosts who are motivated to "meet people" may price their property lower to attract more guests.

Airbnb Data: Hosts' Motivation, Activity, and Retention

We build a Bayesian latent attrition model to capture hosts' retention and activity patterns on Airbnb. It is important to account for host latent attrition because hosts rarely send explicit "exit" signals by removing property profiles or by canceling their host account; they simply become dormant and leave the hosting calendar closed. To calibrate the latent attrition model, we use the nights booked and the price charged for each host's property in each month during the period January 2011 to May 2015 (a total of 53 months).

Method. Following the stream of research on latent attrition models for a noncontractual setting (Fader, Hardie, and Lee 2005; Fader, Hardie, and Shang 2010; Schmittlein, Morrison, and Colombo 1987), we simultaneously model hosts' propensity to churn, the number of booked nights given "alive," and the price per night given a booking. We aggregated hosts' transaction information at a monthly level and incorporated the effect of host motivations as well as unobserved heterogeneity in predicting host activity, retention, and pricing decisions. Our model differs from the traditional Buy 'Till You Die' models in that we allow for flexible heterogeneity in the distributions of the churning probability, booking nights, and price per night (Abe 2009; Padilla and Ascarza 2021; Singh, Borle, and Jain 2009). The model accounts for hosts' unobserved heterogeneity, together with host motivations and the same set of control variables used in the earlier analyses. The key-dependent variables include the number of booked nights per month and the average price per night given a booking. We model monthly booked nights as a zero-inflated Poisson to account for the high proportion of months with no booking, and the price per night as log-normal to account for the skewed distribution of prices. We estimate the model using random-effect Bayesian estimation to allow for unobserved heterogeneity in the model's parameters. See web appendix M.1 for details of the model specification.

Results and Discussion. We report hosts' propensities to churn, the number of booked nights, and price per night using parameters for the posterior mean and 95% central posterior interval. As seen in table 8, hosts motivated to meet people are significantly more engaged in terms of booked nights and have the lowest propensity to churn relative to hosts who do not have that motivation; however, they tend to charge lower prices to entice guests to stay at their property. On the other hand, hosts who are motivated to share beauty tend to have fewer booked nights but charge higher prices for their property. Hosts motivated to earn cash charge lower prices than hosts who do not have that motivation. Lastly, hosts motivated by life circumstances have the highest propensity to churn. The direction and the pattern of the effects for the earn cash motivation are similar to those of the life circumstances motivation.

Table 8 also reveals, as expected, a significant and negative relationship between the proportion of five-star ratings on host likelihood of churn, and a positive and significant relationship with the price charged. This finding supports our conceptualization of a positive relationship between short-term guest satisfaction and long-term host CLV.

Airbnb Data: Hosts' Motivation and Pricing

The results of the latent attrition model (Column 8, table 8) show that hosts with a share beauty motivation are likely to price their property higher, and those with meet people and earn cash motivations are likely to price their property lower than those who do not have these motivations. Those with a life circumstances motivation do not price their property significantly differently from those who do not have that motivation. These findings are consistent with the model-free evidence for the average price charged per night across motivations. This price is the highest among hosts motivated to share beauty ($M_{\text{share beauty}} = \$163.89$),

TABLE 8

BAYESIAN LATENT ATTRITION MODEL ESTIMATES

	Zero-inflated factor Posterior mean (95% CPI)	Booked nights Posterior mean (95% CPI)	Latent churn Posterior mean (95% CPI)	Price per night Posterior mean (95% CPI)
Earn cash	0.02 (−0.04, 0.08)	0.02 (−0.01, 0.04)	0.01 (−0.06, 0.08)	**−0.02** (−0.03, −0.01)
Meet people	**0.10** (0.02, 0.19)	**0.06** (0.04, 0.08)	**−0.12** (−0.19, −0.03)	**−0.06** (−0.07, −0.04)
Share beauty	−0.06 (−0.15, 0.02)	**−0.09** (−0.11, −0.07)	−0.01 (−0.09, 0.07)	**0.05** (0.03, 0.07)
Life circumstances	0.07 (0, 0.13)	**0.05** (0.03, 0.07)	**0.12** (0.05, 0.2)	−0.01 (−0.03, 0)
Guest reviews with five stars (%)	0.01 (−0.16, 0.16)	0.01 (−0.03, 0.04)	**−0.31** (−0.45, −0.17)	**0.19** (0.16, 0.22)
Mean propensities (μ)	0.88 (−0.18, 3.1)	1.72 (0.72, 2.57)	−2.13 (−3.86, −0.73)	5.38 (3.65, 8.68)
Unobserved heterogeneity (σ)	1.57 (1.54, 1.6)	0.56 (0.55, 0.57)	0.41 (0.32, 0.49)	0.43 (0.42, 0.44)

Notes.—The posterior means represent the effects of the covariates on the probability distributions of latent churn, booked nights while alive, and price per night while there is at least one booked night. Numbers in bold indicate that the effects of the variables are significantly different from zero. The table only includes the relevant parameters of interest, see web appendix M.2 for the full table with all the control variables.

the lowest among those motivated to meet people ($M_{\text{meet people}}$ = \$130.25), with those who are motivated to earn cash is in the middle, and close to the share beauty group ($M_{\text{earn cash}}$ = \$158.13). Hosts motivated by life circumstances (\$164.30) charged similarly to those motivated to share beauty. Prices charged differ as predicted between the two intrinsic motivation groups. Given the correlational nature of the pricing differences across motivations in the model-free and latent attrition model, we ran an experiment to test for the causal relationship between motivations and host pricing decisions.

Experiment 2: Pricing Decision

Method. To test the causal influence of motivations on pricing, we manipulated motivations and measured respondents' property pricing decisions, while holding property characteristics constant. MTurk participants ($N = 124$; $M_{\text{age}} = 32.19$, 58.87% male) were randomly assigned to one of three conditions ("earn cash," "share beauty," and "meet people"). Upon starting the experiment, all participants saw images of an apartment in Florida (web appendix N.1). They were then asked to imagine owning the apartment and to consider renting out a room in the apartment on Airbnb. To manipulate motivations, we asked participants to either describe how the Airbnb posting could help them financially (earn cash), how it could help them meet people from around the world (meet people), or how it could help them share beautiful aspects of their homes (share beauty). See web appendix N.2 for details.

All participants spent a minimum of 3 minutes on the description task. A series of manipulation check questions confirmed that our motivation manipulation was successful (see web appendix N.3 for details). Next, all participants responded to two questions of key interest, which were averaged to form the dependent variable ($r = .64$, $p < .001$): (1) "How much would you charge for this private room compared to other similar-quality rooms in the same region?" (7-point scale: 1 = significantly lower, 4 = about the average price, and 7 = significantly higher) and (2) "What would be the appropriate price of your room, given that you have taken the market price into your consideration?" (7-point scale: 1 = significantly lower, 4 = about the market price, 7 = significantly higher). Lastly, participants responded to several open-ended questions about perception/experiences of Airbnb and some demographics questions.

Results and Discussion. One participant who took an extremely long time (40 minutes, above 6 SD of the average of 9 minutes per survey) and two participants who were not in the age range of 18–60 were excluded from analyses. A one-way ANOVA analysis revealed a significant difference in pricing across conditions ($F(2, 118) = 7.67$, $p = .001$). Similar to the model-free results from the Airbnb dataset, hosts who were motivated to share beauty priced their room significantly higher than those who were motivated to earn cash ($M_{\text{share beauty}} = 4.78$, SD = 0.91 vs. $M_{\text{earn cash}} = 4.44$, SD = 0.66; $F(1, 118) = 4.13$, $p = .044$; $d = .43$) and those who were motivated to meet people ($M_{\text{meet people}} = 4.11$, SD = 0.68; $F(1, 118) = 15.34$, $p < .001$; $d = .83$). Hosts who were motivated to meet people charged significantly less than those who were motivated to earn cash ($F(1, 118) = 4.06$, $p = .046$; $d = .49$). These differences across the motivation conditions remained robust ($p < .001$) after controlling for whether participants have had experience hosting on Airbnb ($p = .474$), used Airbnb as guests ($p = .253$), heard of Airbnb before ($p = .004$), and whether they have sublet a property in the past ($p = .520$).[16]

It is possible that the observed differences in price can be explained by participants' inferences about the host's income or the objective value of the property. We,

16 A follow-up pricing experiment revealed similar results with respect to the relationship between host motivation and pricing decisions (for details, see web appendix N.4).

therefore, conducted a post-test (preregistered[17]; detailed procedure can be found in web appendix N.5) where we manipulated the motivations in the same way as in the experiment, and asked participants ($N = 178$, $M_{\text{age}} = 32.76$, 52.8% male) to predict the annual income of the host in the scenario and the property value. Perceived host income and property price were measured using two-questions for each ($r = .78$ for income and $r = .70$ for property price). Participants in the three conditions did not vary in their predictions of the hosts' annual income ($p = .700$) or the value of the property ($p = .272$). Thus, the motivation manipulation does not appear to change inferences about host income or property value, ruling out these factors as potential confounds in the experiment. See web appendix N.6 for detailed results of this supplemental analysis.

Given the differences uncovered between hosts with different motivations in terms of retention, activity, and pricing, we expect the CLV of these different hosts to differ as well. We turn to this analysis next.

CLV Analysis

We use the retention, activity, and pricing results from the latent attrition model to calculate CLV for hosts with different motivations. In the context of the sharing economy, in which customers are microentrepreneurs and firms collect a fixed proportion of the host's revenue, host CLV is proportionally related to firm CLV. We calculated host CLV as the net present value of residual lifetime value (RLV) of a 10-year period (120 months) post the end of the data period (May 2015), plus the realized historical value (HV) during the data period. Note that our CLV calculation is the sales or revenue-based CLV rather than profit CLV as we do not directly observe hosts' costs. Additionally, we do not observe customer acquisition costs (CAC; Gupta et al. 2004). Thus, our CLV calculation is a finite horizon Sales-CLV (excluding CAC).

Method. To calculate the CLV for hosts with different motivations, we need a measure of host retention, the number of booked nights, and the average price per night. Using the Bayesian latent attrition model estimates, we draw from the posteriors of the retention, booked nights, and price per night parameters to calculate the number of months in which hosts stay alive, the number of booked nights given that they are alive, and the average price per night given that they have a booking (Rows 2–4 in table 9). Next, we obtain the HV of each host as the sum of prices of all bookings observed in the transactional data. For each host i, the expected CLV is calculated as follows:

$$E(CLV_i) = HV_i + \sum\nolimits_{t=T_i+1}^{T_i+120} d^{t-T_i} \times PricePerNight_i \times (BookedNights_i | aliveatt) \times P(aliveatt | t > T_i),$$

where $d = 0.99$ is the monthly discount factor.[18] The expected CLV shown in table 9 represents the average host CLV by motivation. For the detailed procedure on CLV calculation using these components, see web appendix O.1.

Results and Discussion. Currently, sharing economy platforms appear to mainly target hosts who have the "earn cash" motivation. In fact, based on our analysis, hosts with the earn cash motivation have the lowest CLV across all motivations. Our results suggest that for the firm to maximize long-term profits, it should also target hosts who are motivated to share beauty as these hosts have a higher CLV than those who are motivated to earn cash. As we suggest in our conceptual model, hosts with intrinsic motivations (share beauty and meet people) are more likely to be retained by the platform than other hosts. If the objective is to acquire hosts who would be most active over time (independent of the revenue they generate), platforms should target hosts who are motivated to meet people.

The results for hosts motivated by life circumstances also deserve discussion. These hosts tend to rent their home for a long-term duration (e.g., "my roommate moved out," "I found a job in another city"). Our dataset reveals that these hosts have a significantly higher proportion of reservations longer than a month (12.44%) compared to other hosts (earn cash: 10.80%, meet people: 10.37%, share beauty: 10.68%). However, the life circumstances motivation is temporary and is likely to change as a host's life situation changes; hence this motivation is less actionable for Airbnb. We also note that calculating CLV ignoring host motivation (table 9, Row 6) can lead the firm to miss out on attractive segmenting and targeting opportunities.

Related to the reported relationship between guest satisfaction and the latent attrition model outcomes, we also find that hosts with 90% of five-star guest ratings (high level of guest satisfaction) have a significantly higher CLV ($30,664.48) compared to hosts with 60% of five-star guest ratings (low level of guest satisfaction; $28,773.95). This result demonstrates the link between short-term guest satisfaction and long-term host CLV (see web appendix O.2 for details of this analysis).

17 https://aspredicted.org/sh3ex.pdf

18 The discount factor represents the weighted average cost of capital (WACC) of the firm (McCarthy, Fader, and Hardie 2017). For privately held companies, it is common to estimate that WACC is equal to the average for the sector that the company operates in. In 2015, WACC was approximately 13%, which is largely in agreement with a 0.99 monthly discount rate (Damodaran 2021).

TABLE 9

HOST CLV BY MOTIVATION

	Earn cash	Meet people	Share beauty	Life circumstances
Expected number of months alive	26.89	33.91	33.05	28.40
Expected number of booked nights when alive	9.83	9.86	9.21	10.17
Expected price per night	154.91	126.85	159.57	159.58
Expected CLV	28,579.67	30,567.08	31,867.42	31,166.86
Expected CLV without considering motivations	29,878.22			

Notes: The inferred number of booked nights and price per night was multiplied by a factor that was randomly drawn from Uniform (0.5, 1.5) to disguise the actual values for confidentiality reasons. The calculated expected CLV is also scaled, accordingly.

GENERAL DISCUSSION

This article demonstrates how firms can leverage machine learning tools to extract consumer motivations from textual data, and then assesses the downstream consequences of motivations on outcomes such as host engagement, activity, retention, pricing, and guest satisfaction. We find that hosts with intrinsic motivations (share beauty and meet people) tend to be more engaged than those with extrinsic motivations (earn cash) when listing their property on Airbnb, as seen in the amount of text and photos they post. Reflecting the host's engagement in promoting their property, guests who stayed at these properties reported a higher level of satisfaction. Hosts motivated by sharing beauty and meeting people were also more likely to stay active on the platform in the long term. Finally, hosts with the sharing beauty motivation have the highest CLV whereas hosts with the earning cash motivation have the lowest CLV. By documenting the effect of a distal psychological variable on firm-relevant metrics, our work contributes to a small but growing body of research that bridges the consumer behavior and marketing science disciplines (Bell and Lattin 2000; Berger and Milkman 2012; Berger et al. 2020; Humphreys and Wang 2018; Kivetz, Netzer, and Srinivasan 2004).

We found that hosts may participate in the sharing economy for multiple reasons (web appendix E.1). One may raise the concern that all hosts actually desire only to earn cash and that some hosts are tactfully disguising this true motivation by expressing their desire to socialize (meet people) or to be altruistic (share beauty). Findings from our experiments that manipulate motivations and replicate key findings from the secondary data should help rule out this concern. This alternative account is also rendered unlikely by our findings of behavioral differences that are consistent with predictions from self-determination theory. One may question why the monetary motive (extrinsic) does not crowd out social, intrinsic motives as research has previously found (Deci et al. 2001; Heyman and Ariely 2004). In the context of the sharing economy, the monetary transaction can be viewed as an assurance structure that facilitates social exchanges between hosts and guests (Lampinen and Cheshire 2016). The "earn cash" motivation need not crowd out the joy of hosting and can even be a facilitator of social benefits. One caveat that we acknowledge is that the factors we control in our analyses such as income and property value could well be ecologically related to the behavioral outcomes we study such as engagement and pricing. Our findings show that motivation plays are role even after controlling for these effects.

This research makes a methodological contribution to the consumer behavior literature by providing practical tools that academics and practitioners can use to extract latent psychological traits such as consumer motivations (Berger et al. 2020; Matz and Netzer 2017). We document the use of two machine learning approaches (multilabel classification and Poisson factorization) to extract motivations from open-ended survey responses. Our approach can be used to extract motivations from other existing textual data, such as customer reviews on a firm's webpage, commercial websites (e.g., Amazon, Yelp, BestBuy), or consumer blogs. Our approach can also be applied in other business and societal contexts. Real estate analysts (e.g., Zillow) can use simple open-ended surveys to identify reasons for the influx/exit of home-buyers in a neighborhood (e.g., good school district, nature, privacy, noise level). Headhunting agencies can identify individuals' motivations underlying one's job search. Educators in K-12 schools can use students' verbal responses to a short essay prompt to uncover different motivations for learning—an approach that overcomes young students' potential lack of comprehension of survey questions and response scales. While the traditional approach of using closed-ended scales to measure motivations is useful for theory testing, it has limited applicability in the real world where motivations are often unknown and may coexist.

Our research also makes a substantive contribution by demonstrating that consumer motivations can have real and long-lasting downstream consequences on consumer behaviors that are relevant to firms. Such analyses can help firms uncover which consumer segments are likely to engage closely with the firm over time and provide the greatest lifetime value and shed light on how best to target these segments. We demonstrate these effects in the unique context of the sharing economy where the company serves as a platform and customers interact with each other to generate revenue (Botsman and Roo 2010; Heinrichs 2013; Lamberton 2016).

Our findings provide practical implications for Airbnb. Acquiring and retaining hosts is of major importance to the financial stability and growth of sharing economy platforms. Given heterogeneity in motivations, the company could diversify its approach to acquiring potential customers by targeting hosts who are motivated to meet others or to share beauty because they tend to be more engaged with guests and the firm. Hosts with these motivations also provide higher (sales-) CLV than monetary motivated hosts. Contrary to these implications, Airbnb's acquisition advertisements tend to focus on the economic benefits and often target monetarily motivated hosts. This focus could indicate that the CAC for earn cash-motivated hosts are higher than the CAC for hosts with other motivations, who join without such targeted advertising. This view suggests that the difference in CLV across motivations after accounting for CAC could be even larger than the CLV reported in the article. A large proportion of the firm's profit may actually be generated from a smaller proportion of the intrinsically motivated hosts (McCarthy and Winer 2019).

On the other hand, it is possible that the firm does not target hosts with other motivations because they have a higher CAC or more simply, because they do not have much insight into the prevalence of these motivations. Future research is needed to examine the firm's ability and cost to target different motivation segments. Research on the sharing economy could also explore how to use text extracted from other sources (e.g., social media posts, product reviews) to reliably extract motivations including potentially using automated machine-learning methods (Moe, Netzer, and Schweidel 2017). The three motivations we found are likely to be unique to Airbnb. For example, Uber drivers are likely to be motivated to earn cash and to meet people, but they are less likely to be motivated to share the beautiful interior of their cars. This speculation can be tested using the approach identified in our article.

In sum, this research demonstrates the value of extracting motivations from textual data and exploring the downstream consequences of such motivations using transactional data. Future research can build on this work by extracting "soft" latent dimensions of consumer psychology and incorporating them into marketing analytics models to investigate their real-world consequential impact.

DATA COLLECTION INFORMATION

The last author extracted proprietary data from Airbnb upon approval from the company. The first and third authors jointly analyzed the data under the supervision of the second and fourth authors. Additionally, the first author conducted the first experiment in July 2019 using Prolific platform, and the second experiment in April 2016 using MTurk platform. Both experiments were analyzed by the first author under the supervision of the second author. The last author provided insights from the company during the process of analyses. All notes, images, and data are currently stored in a Dropbox folder under the management of the first author.

REFERENCES

Abe, Makoto (2009), "Counting Your Customers' One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model," *Marketing Science*, 28 (3), 541–53.

Airbnb (2020), "Airbnb," https://news.airbnb.com/fast-facts/ (last accessed 05 June 2021).

—— (2021), "What Are Airbnb Service Fees? Airbnb Help Center," *Airbnb*, https://www.airbnb.com/help/article/1857/what-are-airbnb-service-fees (last accessed 05 June 2021).

Aspara, Jaakko and Kristina Wittkowski (2019), "Sharing-Dominant Logic? Quantifying the Association between Consumer Intelligence and Choice of Social Access Modes," *Journal of Consumer Research*, 46 (2), 201–22.

Bauer, Karen W. (2004), "Conducting Longitudinal Studies," *New Directions for Institutional Research*, 2004 (121), 75–90.

Bell, David R. and James M. Lattin (2000), "Looking for Loss Aversion in Scanner Panel Data: The Confounding Effect of Price Response Heterogeneity," *Marketing Science*, 19 (2), 185–200.

Benkler, Yochai. (2004), "Sharing Nicely: On Shareable Goods and the Emergence of Sharing as a Modality of Economic Production," http://www.yalelawjournal.org/essay/sharing-nicely-on-shareable-goods-and-the-emergence-of-sharing-as-a-modality-of-economic-production (last accessed 05 June 2021).

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.

Berger, Jonah and Katherine L. Milkman (2012), "What Makes Online Content Viral?," *Journal of Marketing Research*, 49 (2), 192–205.

Bernard, H. Russell and Harvey Russell Bernard. (2013), *Social Research Methods: Qualitative and Quantitative Approaches*, SAGE Publications. https://books.google.com/books?id=7sZHuhyzBNQC (last accessed 05 June 2021).

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3 (January), 993–1022.

Blue, George M. and Minah Harun (2003), "Hospitality Language as a Professional Skill," *English for Specific Purposes*, 22 (1), 73–91.

Bogin, Alexander, William Doerner, and William Larson (2019), "Local House Price Dynamics: New Indices and Stylized Facts," *Real Estate Economics*, 47 (2), 365–98.

Borucki, Chester C. and Michael J. Burke (1999), "An Examination of Service-Related Antecedents to Retail Store Performance," *Journal of Organizational Behavior*, 20 (6), 943–62.

Botsman, Rachel and Rogers Roo (2010), "Beyond Zipcar: Collaborative Consumption," *Harvard Business Review*, 88 (10), 30.

Buhalis, Dimitrios, Luisa Andreu, and Juergen Gnoth (2020), "The Dark Side of the Sharing Economy: Balancing Value Co-Creation and Value Co-Destruction," *Psychology & Marketing*, 37 (5), 689–704.

Canny, John (2004), "GaP: A Factor Model for Discrete Data," in *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA: ACM, 122–29.

Census (2021), "Survey/Program: American Community Survey TableID: S1903," *Census.gov*, https://www.census.gov/en.html (last accessed 05 June 2021).

Chen, Yubo and Liantao (Tarry) Wang (2019), "Commentary: Marketing and the Sharing Economy: Digital Economy and Emerging Market Challenges," *Journal of Marketing*, 83 (5), 28–31.

Cho, Cecile K. and Gita Venkataramani Johar (2011), "Attaining Satisfaction," *Journal of Consumer Research*, 38 (4), 622–31.

Choy, Looi Theam (2014), "The Strengths and Weaknesses of Research Methodology: Comparison and Complimentary between Qualitative and Quantitative Approaches," *IOSR Journal of Humanities and Social Science*, 19 (4), 99–104.

Damodaran, Aswath (2021), "Total Betas by Sector Dataset," http://pages.stern.nyu.edu/~adamodar/New_Home_Page/dataarchived.html (last accessed 05 June 2021).

Deci, Edward L., Richard Koestner, and Richard M. Ryan (2001), "Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again," *Review of Educational Research*, 71 (1), 1–27.

Deci, Edward L. and Richard M. Ryan (1985), "The General Causality Orientations Scale: Self-Determination in Personality," *Journal of Research in Personality*, 19 (2), 109–34.

Dommer Sara, Vanitha Swaminathan, Loughran, and Rohini Ahluwalia (2013), "Using Differentiated Brands to Deflect Exclusion and Protect Inclusion: The Moderating Role of Self-Esteem on Attachment to Differentiated Brands," *Journal of Consumer Research*, 40 (4), 657–75.

Drèze, Xavier, and Joseph C. Nunes (2009), "Feeling Superior: The Impact of Loyalty Program Structure on Consumers' Perceptions of Status," *Journal of Consumer Research*, 35 (6), 890–905.

Dudwick, Nora, Kathleen Kuehnast, Veronica Nyhan Jones, and Michael Woolcock (2006), "Analyzing Social Capital in Context: A Guide to Using Qualitative Methods and Data," http://pages.stern.nyu.edu/~adamodar/New_Home_Page/dataarchived.html (last accessed 05 June 2021).

Eckhardt, Giana M., Mark B. Houston, Baojun Jiang, Cait Lamberton, Aric Rindfleisch, and Georgios Zervas (2019), "Marketing in the Sharing Economy," *Journal of Marketing*, 83 (5), 5–27.

Fader, Peter S., Bruce G.S. Hardie, and Ka Lok Lee (2005), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, 42 (4), 415–30.

Fader, Peter S., Bruce G.S. Hardie, and Jen Shang (2010), "Customer-Base Analysis in a Discrete-Time Noncontractual Setting," *Marketing Science*, 29 (6), 1086–108.

Geissinger, Andrea, Christofer Laurell, and Christian Sandström (2020), "Digital Disruption beyond Uber and Airbnb—Tracking the Long Tail of the Sharing Economy," *Technological Forecasting and Social Change*, 155, 119323.

Genesove, David and Christopher Mayer (2001), "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *The Quarterly Journal of Economics*, 116 (4), 1233–60.

Gopalan, Prem, Jake M., Hofman, and David M. Blei (2013), "Scalable Recommendation with Poisson Factorization," *arXiv:1311.1704 [cs, stat]*, http://arxiv.org/abs/1311.1704.

Government of Canada, Statistics Canada (2021), "Statistics Canada: Canada's National Statistical Agency," https://www.statcan.gc.ca/eng/start (last accessed 05 June 2021).

Grant, Adam M. (2008), "Does Intrinsic Motivation Fuel the Prosocial Fire? Motivational Synergy in Predicting Persistence, Performance, and Productivity," *The Journal of Applied Psychology*, 93 (1), 48–58.

Gruman, Jamie A. and Alan M. Saks (2011), "Performance Management and Employee Engagement," *Human Resource Management Review*, 21 (2), 123–36.

Gundersen, Marit G., Morten Heide, and Uff H. Olsson (1996), "Hotel Guest Satisfaction among Business Travelers: What Are the Important Factors?," *Cornell Hotel and Restaurant Administration Quarterly*, 37 (2), 72–81.

Gupta, Sunil and Donald R. Lehmann (2006), "Customer Lifetime Value and Firm Valuation," *Journal of Relationship Marketing*, 5 (2–3), 87–110.

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

Hamari, Juho, Mimmi Sjöklint, and Antti Ukkonen (2016), "The Sharing Economy: Why People Participate in Collaborative Consumption," *Journal of the Association for Information Science and Technology*, 67 (9), 2047–59.

Harter, James K., Frank L. Schmidt, and Theodore L. Hayes (2002), "Business-Unit-Level Relationship between Employee Satisfaction, Employee Engagement, and Business Outcomes: A Meta-Analysis," *The Journal of Applied Psychology*, 87 (2), 268–79.

Heinrichs, Harald (2013), "Sharing Economy: A Potential New Pathway to Sustainability," *Gaia - Ecological Perspectives for Science and Society*, 22 (4), 228–31.

Herbst, Kenneth C., Eli J. Finkel, David Allan, and Gráinne M. Fitzsimons (2012), "On the Dangers of Pulling a Fast One: Advertisement Disclaimer Speed, Brand Trust, and Purchase Intention," *Journal of Consumer Research*, 38 (5), 909–19.

Heyman, James and Dan Ariely (2004), "Effort for Payment: A Tale of Two Markets," *Psychological Science*, 15 (11), 787–93.

Higgins, E. Tory (1998), "Promotion and Prevention: Regulatory Focus as a Motivational Principle," *Advances in Experimental Social Psychology*, 30, 1–46.

Hong, Jiewen and Angela Y. Lee (2008), "Be Fit and Be Strong: Mastering Self- Regulation through Regulatory Fit," *Journal of Consumer Research*, 34 (5), 682–95.

Humphreys, Ashlee and Rebecca Jen-Hui Wang (2018), “Automated Text Analysis for Consumer Research,” *Journal of Consumer Research*, 44 (6), 1274–306.

Jayaweera, Thushel (2015), “Impact of Work Environmental Factors on Job Performance, Mediating Role of Work Motivation: A Study of Hotel Sector in England,” *International Journal of Business and Management*, 10 (3), 271–8.

Karatepe, Osman M. (2013), “High-Performance Work Practices and Hotel Employee Performance: The Mediation of Work Engagement,” *International Journal of Hospitality Management*, 32, 132–40.

Karatepe, Osman M. and Georgiana Karadas (2015), “Do Psychological Capital and Work Engagement Foster Frontline Employees’ Satisfaction? A Study in the Hotel Industry,” *International Journal of Contemporary Hospitality Management*, 27 (6), 1254–78.

Kim, Hyeongmin (Christian) (2013), “Situational Materialism: How Entering Lotteries May Undermine Self-Control,” *Journal of Consumer Research*, 40 (4), 759–72.

Kivetz, Ran, Oded Netzer, and V. Srinivasan (2004), “Alternative Models for Capturing the Compromise Effect,” *Journal of Marketing Research*, 41 (3), 237–57.

Kivetz, Ran, Oleg Urminsky, and Yuhuang Zheng (2006), “The Goal-Gradient Hypothesis Resurrected: Purchase Acceleration, Illusionary Goal Progress, and Customer Retention,” *Journal of Marketing Research*, 43 (1), 39–58.

Kumar, V., Avishek Lahiri, and Orhan Bahadir Dogan (2018), “A Strategic Framework for a Profitable Business Model in the Sharing Economy,” *Industrial Marketing Management*, 69, 147–60.

Lakhani, Karim R. and Robert G. Wolf. (2003), *Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects*, SSRN Scholarly Paper ID 443040, Rochester, NY: Social Science Research Network.

Lamberton, Cait (2016), “Collaborative Consumption: A Goal-Based Framework,” *Current Opinion in Psychology*, 10, 55–59.

Lamberton, Cait Poynor (2015), *Consumer Sharing: Collaborative Consumption, from Theoretical Roots to New Opportunities*, SSRN Scholarly Paper ID 2726285, Rochester, NY: Social Science Research Network, http://papers.ssrn.com/abstract=2726285 (last accessed 05 June 2021).

Lampinen, Airi and Coye Cheshire (2016), “Hosting via Airbnb: Motivations and Financial Assurances in Monetized Network Hospitality,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, 1669–80, http://doi.acm.org/10.1145/2858036.2858092.

Li, Jun, Antonio Moreno, and Dennis J. Zhang (2016), Pros vs Joes: Agent Pricing Behavior in the Sharing Economy, SSRN Scholarly Paper ID 2708279, Rochester, NY: Social Science Research Network, https://papers.ssrn.com/abstract=2708279.

Liu, Yang, Roy Chen, Yan Chen, Qiaozhu Mei, and Suzy Salib. (2012), “‘I Loan Because…’: Understanding Motivations for Pro-Social Lending,” in *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, New York, NY, USA: ACM, 503–12, http://doi.acm.org/10.1145/2124295.2124356.

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

Maylett, Tracy and Paul Warner (2014), *MAGIC: Five Keys to Unlock the Power of Employee Engagement - Tracy Maylett, Paul Warner - Google Books*, Austin, TX: Greenleaf Book Group Press.

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.

McCarthy, Daniel M. and Russell S. Winer (2019), “The Pareto Rule in Marketing Revisited: Is It 80/20 or 70/20?,” *Marketing Letters*, 30 (2), 139–50.

Moe, Wendy W., Oded Netzer, and David A. Schweidel (2017), “Social Media Analytics,” in *Handbook of Marketing Decision Models*, ed. Wierenga Berend and Ralf van der Lans, Cham: Springer International Publishing, 483–504, 10.1007/978-3-319-56941-3_16.

Moorman, Christine, Harald J. van Heerde, C. Page Moreau, and Robert W. Palmatier (2019), “Challenging the Boundaries of Marketing,” *Journal of Marketing*, 83 (5), 1–4.

Morstatter, Fred and Liu Huan (2017), “In Search of Coherence and Consensus: Measuring the Interpretability of Statistical Topics,” *Journal of Machine Learning Research*, 18 (1), 6177–208.

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.

Nov, Oded (2007), “What Motivates Wikipedians?,” *Communications of the ACM*, 50 (11), 60–64.

Ou, Shiyan, Christopher Soo-Guan Khoo, and Dion H. Goh (2008), “Design and Development of a Concept-Based Multi-Document Summarization System for Research Abstracts,” *Journal of Information Science*, 34 (3), 308–26.

Ozanne, Lucie K. and Paul W. Ballantine (2010), “Sharing as a Form of Anti-Consumption? An Examination of Toy Library Users,” *Journal of Consumer Behaviour*, 9 (6), 485–98.

Packard, Grant and Jonah Berger (2017), “How Language Shapes Word of Mouth’s Impact,” *Journal of Marketing Research*, 54 (4), 572–88.

Padilla, Nicolas and Eva Ascarza (2021), *Overcoming the Cold Start Problem of CRM Using a Probabilistic Machine Learning Approach*, SSRN Scholarly Paper ID 2933291, Rochester, NY: Social Science Research Network, https://papers.ssrn.com/abstract=2933291.

Pennebaker, James W., Ryan L. Boyd, Kayla Jordan, and Kate Blackburn (2015), *The Development and Psychometric Properties of LIWC2015*, Austin, TX: University of Texas at Austin.

Permut, Stephanie, Matthew Fisher, and Daniel M. Oppenheimer (2019), “TaskMaster: A Tool for Determining When Subjects Are on Task,” *Advances in Methods and Practices in Psychological Science*, 2 (2), 188–96.

Pieters, Rik (2017), “Meaningful Mediation Analysis: Plausible Causal Inference and Informative Communication,” *Journal of Consumer Research*, 44 (3), 692–716.

Poynor, Cait and Kelly L. Haws (2009), “Lines in the Sand: The Role of Motivated Categorization in the Pursuit of Self-Control Goals,” *Journal of Consumer Research*, 35 (5), 772–87.

Queirós, André, Faria Daniel, and Almeida Fernando (2017), "Strengths and Limitations of Qualitative and Quantitative Research Methods," *European Journal of Education Studies*, 3 (9), 369–87. DOI: 10.5281/zenodo.887089.

Roberts, Carl W. (2020), *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*, New York: Routledge. https://www.taylorfrancis.com/books/edit/10.4324/9781003064060/text-analysis-social-sciences-carl-roberts

Ryan, Richard M. and Edward L. Deci (2000), "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions," *Contemporary Educational Psychology*, 25 (1), 54–67.

—— (2020), "Intrinsic and Extrinsic Motivation from a Self-Determination Theory Perspective: Definitions, Theory, Practices, and Future Directions," *Contemporary Educational Psychology*, 61, 101860.

Saks, Alan M. (2006), "Antecedents and Consequences of Employee Engagement," *Journal of Managerial Psychology*, 21 (7), 600–19.

Salton, Gerard and Christopher Buckley (1988), "Term-Weighting Approaches in Automatic Text Retrieval," *Information Processing & Management*, 24 (5), 513–23.

Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who-Are They and What Will They Do Next?," *Management Science*, 33 (1), 1–24.

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.

Sievert, Carson and Kenneth Shirley (2014), "LDAvis: A Method for Visualizing and Interpreting Topics," in *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, Baltimore, MD: Association for Computational Linguistics, 63–70, https://www.aclweb.org/anthology/W14-3110 (last accessed 05 June 2021).

Singh, Siddharth S., Sharad Borle, and Dipak C. Jain (2009), "A Generalized Framework for Estimating Customer Lifetime Value When Customer Lifetimes Are Not Observed," *Quantitative Marketing and Economics*, 7 (2), 181–205.

So, Kevin Kam Fung, Hyunsu Kim, and Haemoon Oh (2020), "What Makes Airbnb Experiences Enjoyable? The Effects of Environmental Stimuli on Perceived Enjoyment and Repurchase Intention," *Journal of Travel Research*, 60 (5), 1018–38.

Soman, Dilip and Amar Cheema (2004), "When Goals Are Counterproductive: The Effects of Violation of a Behavioral Goal on Subsequent Performance," *Journal of Consumer Research*, 31 (1), 52–62.

Spinelli, Michael and George Canavos (2000), "Investigating the Relationhip between Employee Satisfaction and Guest Satisfaction—Michael A. Spinelli, George C. Canavos, 2000," *Cornell Hotel and Restaurant Administration Quarterly*, 41 (6), 29–33.

Tabcum, Sarote (2019), "The Sharing Economy Is Still Growing, and Businesses Should Take Note," *Forbes*, https://www.forbes.com/sites/forbeslacouncil/2019/03/04/the-sharing-economy-is-still-growing-and-businesses-should-take-note/ (last accessed 05 June 2021).

Toubia, Olivier and Oded Netzer (2017), "Idea Generation, Creativity, and Prototypicality," *Marketing Science*, 36 (1), 1–20.

Tsaur, Sheng-Hshiung, Fu-Sung Hsu, and Hsin Lin (2019), "Workplace Fun and Work Engagement in Tourism and Hospitality: The Role of Psychological Capital," *International Journal of Hospitality Management*, 81, 131–40.

Tsoumakas, Grigorious and Ioannis Katakis (2007), "Multi-label classification: An overview," *International Journal of Data Warehousing and Mining (IJDWM)*, 3 (3), 1–13.

van der Heijden, Hans (2004), "User Acceptance of Hedonic Information Systems," *MIS Quarterly*, 28 (4), 695–704.

Viglia, Giampaolo (2020), "The Sharing Economy: Psychological Mechanisms That Affect Collaborative Consumption," *Psychology & Marketing*, 37 (5), 627–29.

Yi, Jisu, Gao Yuan, and Changsok Yoo (2020), "The Effect of the Perceived Risk on the Adoption of the Sharing Economy in the Tourism Industry: The Case of Airbnb," *Information Processing & Management*, 57 (1), 102108.

Zervas, Georgios, Davide Proserpio, and John W. Byers (2017), "The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry," *Journal of Marketing Research*, 54 (5), 687–705.

Zhang, Zhihua, Rachel J.C. Chen, Lee D. Han, and Lu Yang (2017), "Key Factors Affecting the Price of Airbnb Listings: A Geographically Weighted Approach," *Sustainability*, 9 (9), 1635.