



Understanding the impact of host facial characteristics on Airbnb pricing: Integrating facial image analytics into tourism research

Stuart J. Barnes^{a,*}, Samuel Nathan Kirshner^b

^a CODA Research Centre, King's Business School, King's College London, Bush House, 30 Aldwych, London, WC2B 4BG, United Kingdom

^b School of Information Systems and Technology Management, UNSW Business School, University of New South Wales, Room 2114, Quadrangle Building, Sydney, 2052, Australia

ARTICLE INFO

Keywords:

Host
Face
Image analytics
Convolutional neural network
Trust
Attractiveness
Price
Deep learning

ABSTRACT

People rapidly and subconsciously process information from facial images. On sharing economy platforms, facial cues can provide a useful supplement to other information provided by reputation systems. Previous small-scale, rater-informed studies examining trust and attractiveness based on facial features on Airbnb found mixed support for impacts on pricing. We re-examine their impact using deep learning to classify host faces for an extensive data set of Airbnb accommodation in 10 US cities ($n = 78,215$). Together, trust and attractiveness contribute to almost a 5% increase in prices for Airbnb accommodation. We also test Gray's theory of motivation via the examination of pricing for different types of accommodation, finding that trust is more important in situations of smaller accommodation shared with strangers. The paper concludes with limitations and implications for research and practice.

1. Introduction

The sharing economy market is estimated to reach \$40.2 billion in revenues by 2022, up from \$18.6 billion in 2017, driven mainly by transport and accommodation (Juniper Research, 2017). As the largest provider of accommodation, Airbnb had more than 6 million listings and 150 million users in June 2019 and projected revenues of \$8.5 billion in 2020 (iPropertyManagement, 2019). Future revenue growth will depend on the further development of consumer demand alongside improving revenues for existing accommodation.

Reputation systems used on sharing economy platforms provide valuable tools to encourage and reduce the risk of participation. However, reputation systems have limitations and are typically supplemented with other items, such as host images, which may be rapidly processed by participants to provide information like a host's perceived trustworthiness. This section explains the motivation for this study, demonstrating the mixed results and limitations of previous studies that examine the impact of host facial characteristics on Airbnb pricing. We further explain the contribution of this study, including a large-scale deep learning approach to test previous models, and an extension of prior research via hypotheses to examine the scope of the impact of trust when a property is shared and to test the non-linear effects of the

number of visitor reviews on price.

1.1. Motivation for the study

Transactions in online peer-to-peer (P2P) markets contain risks for buyers pertaining to the accuracy of product/service descriptions, the possibility that the sellers will fail to fulfil their contractual obligations in prepaid transactions, and the possible privacy breach of personal and financial information disclosed by buyers to initiate a transaction (Pavlou, Liang, & Xue, 2007). Thus, the success of online marketplaces is contingent on the buyer's trust that the seller is acting in good faith in terms of information disclosure and their intention to fulfil contractual obligations. To mitigate buyers' risk, online platforms have established reputation systems based on participants' past performance and other mechanisms to regulate the transparency of information and facilitate trust (Luca, 2017). Reputation systems are usually based on the quantity and content of reviews from counterparties in past transactions and provide signals on transaction quality, whereas other platform design-based mechanisms facilitate trust by providing cues on the characteristics of participants (Dellarocas, 2003; Luca, 2017).

Early research on reputation mechanisms in P2P e-commerce found a strong link between sellers' reputation and buyers' willingness to pay (e.

* Corresponding author.

E-mail addresses: stuart.barnes@kcl.ac.uk (S.J. Barnes), s.kirshner@unsw.edu.au (S.N. Kirshner).

<https://doi.org/10.1016/j.tourman.2020.104235>

Received 22 July 2019; Received in revised form 18 September 2020; Accepted 23 September 2020

Available online 29 September 2020

0261-5177/© 2020 Elsevier Ltd. All rights reserved.

g., Houser & Wooders, 2006; Resnick; Zeckhauser, Swanson, & Lockwood, 2006). However, recent research questions the effectiveness of traditional reputation mechanisms. Bolton, Greiner, and Ockenfels (2013) find strong evidence that reciprocity in eBay's review system reduces the informativeness of feedback. Sellers strategically delay reviewing buyers until after receiving their review and reciprocate positive and negative feedback which discourages future negative reviews. From a larger internal dataset, Nosko and Tadelis (2015) find that the mean percentage of positive reviews on eBay is 99.3% and that negative experiences result in formal complaints to customer service rather than negative reviews, further skewing the positive reputation bias.

Sharing economy platforms have additional risks for buyers and sellers beyond traditional online P2P markets selling goods and services. Sellers face risks to their property, whereas buyers' safety may be compromised when in unfamiliar environments. For example, Uber drivers and passengers have been assaulted and robbed (Feeney, 2015). Likewise, Airbnb hosts have experienced property damage and theft (ter Huurne, 2017), while guests have had their security and privacy compromised (Zhang, Yan, & Zhang, 2018). Recently, multiple accounts have surfaced of hosts live streaming video feeds of their guests. There are additional concerns with compatibility for ridesharing and co-occupation of accommodation. This is particularly true of Airbnb, where many accommodations for rent are shared with hosts. These problems are compounded by 'gray areas' of the sharing economy regarding liabilities and legal protection when buyers or sellers incur damages or injuries (Ranchordás, 2015).

Unfortunately, reputation systems in the sharing economy may not reliably ensure the transparency of information or alert participants to potential dangers. These systems tend to have similar biases associated with rating systems in traditional P2P markets which positively skew ratings. In the case of Airbnb, Zervas, Proserpio, and Byers (2015) show that the reciprocal rating system, where participants in a transaction rate one another, produces more positive ratings on the same properties which are cross listed on other platforms. As a result, 95% of Airbnb properties have an average rating above 4.5 (out of 5) and virtually no properties rate below 3.5. Bridges and Vásquez (2018) find almost all written reviews are brief and highly positive (e.g., "Great guests! Would love to host again!"). Moreover, the continuous expansion of participation in sharing marketplaces implies that new participants often have limited (if any) reviews. Therefore, inferring trust from reputation systems is challenging for both hosts and guests (Zhang et al., 2018).

Given the limitations of reputation systems, sharing economy platforms are designed to provide users with additional information and assurances on market participants (Luca, 2017). Verification of the user's identity and profile photos are standard features of sharing platforms since they are a powerful mechanism for creating trust between P2P users (Bente, Baptist, & Leuschner, 2012; Jaeger, Slegers, Evans, Stel, & van Beest, 2018; Luca, 2017; Teubner, Adam, Camacho, & Hassanein, 2014). Hosts should pay attention to the image conveyed not just by the reputation system, but also from additional sources of information provided to guests. Such information gives guests additional cues on aspects of the host that may influence value.

High-quality profile photos that make the host appear trustworthy and attractive are more desirable to guests and, thus, are more likely to command a price premium. This is a key focus of our study. Ert, Fleischer, and Magen (2016) investigated the impact of reputation scores and perceived trustworthiness and attractiveness of host photos on Airbnb prices. Ert et al. (2016) hired Mechanical Turk (MTurk) workers to independently evaluate the trustworthiness and attractiveness of hosts for 175 Airbnb properties in Stockholm, Sweden. Controlling for the attributes of the accommodations, they found an absence of variation in review scores and that the perceived trustworthiness of hosts (based on their photo) significantly impacted prices. They found less support for a relationship between attractiveness and listing prices. In a separate analysis, Ert et al. (2016) manipulated reputation scores to

increase variability. While reputation had a modest impact on prices, they found that perceived trustworthiness from photos still had the largest effect on price premiums. Jaeger et al. (2018) attempted to corroborate the results of Ert et al. (2016) by controlling for additional variables and extending the size of the study to 1020 listings in New York City. In contrast to Ert et al. (2016), they found that attractiveness, not trustworthiness, led to higher prices. Together, these studies suggest that guests' impressions of profile photos influence pricing, but it is unclear whether the price premium is driven by perceived trustworthiness or attractiveness.

There are several limitations to the study designs of Ert et al. (2016) and (Jaeger et al. (2019)) which may account for the variations in findings. First, the MTurk workers who performed the ratings in both studies were self-selected. Second, the scope of the two studies was resource-constrained due to the use of MTurk, which inherently limits the scale of a study (although it provides a wider demographic of participants than traditional procedures using undergraduate students as test subjects). Finally, previous studies do not examine the interaction of perceived trust according to the types of accommodation, particularly the degree of sharing and size of accommodation.

1.2. Organization

The rest of the article is organized as follows: In the next section, we review related literature to build our hypotheses. In the third section, we introduce a research framework for integrating facial image classification into social science studies. In section four, we present and discuss our empirical results, comparing the results with the salient literature. In the final section, we present our conclusions, study limitations, and the implications of our work for research and practice.

2. Hypothesis development

In advertising, physical attractiveness, and perceived friendliness enhance credibility (Joseph, 1982). The same principle applies to sharing platforms where these attributes in a participant's photos can foster the impression of trustworthiness by signalling kindness, honesty, and politeness (Bente et al., 2012; Zhang et al., 2018). The combination of a photo and profile also promotes trustworthiness by reducing identity ambiguity (Bente et al., 2012; Guttentag, 2015) and can lessen the sense of social distance between users, which impacts the perception of risk (Teubner et al., 2014). Thus, information on the host is a central feature for the platform (Zekanovic-Korona & Grzunov, 2014), and a guest may choose where to stay based on a host profile picture (Luca, 2017).

Facial cues provide an important source of information. Observers often make inferences on people's general disposition and specific trait impressions (Hassin & Trope, 2000). Willis and Todorov (2006) conducted a series of experiments manipulating the time that subjects were exposed to photos of faces to understand the speed of trait inferences regarding attractiveness, likeability, trustworthiness, competence, and aggressiveness. Across all five traits, they found that judgements formed after an exposure time of 100 ms correlated with judgements made without time constraints and that judgements made after 100 ms did not statistically increase with greater exposure time. Willis and Todorov (2006) hypothesized that attractiveness would have the highest correlation with evaluations made without time constraints since it is a property of facial appearance. While likeability, competence, and aggressiveness had lower correlations compared to attractiveness, they found that judgements of trustworthiness had the highest correlation. Willis and Todorov (2006) argued that this finding was unsurprising as detecting trustworthiness is necessary for human survival, and judging trustworthiness may be an automatic processing activity in the amygdala, which is responsible for detecting dangerous stimuli. In follow-up studies, Todorov, Pakrashi, and Oosterhof (2009) replicated their earlier findings with shorter exposures showing that trustworthiness

judgements can occur within 33 ms, and Engell, Haxby, and Todorov (2007) provided further support that the amygdala automatically evaluates trustworthiness based on facial cues that are commonly perceived as signals of trustworthiness.

A common assumption is that people deliberately evaluate information before making important social or economic decisions. However, unreflective traits rapidly inferred from photos can significantly influence how subsequent information is processed, which impact people's decisions. For example, Todorov, Mandisodza, Goren, and Hall (2005) showed that spontaneous inferences of competence made after a 1-s exposure to political candidates correctly predicted almost 70% of the winners in the 2004 congressional election and were correlated to the margin of victory. Van't Wout and Sanfey (2008) examined the impact of perceived trustworthiness based on facial features from photos and the extent that participants cooperated with strangers in an economic game of trust. They found that people made greater investments when paired with people who were rated as being trustworthy, providing evidence that automatic assessments of trust influence trust behaviour. Several studies utilizing trust games have found similar results providing robust support for the influence of perceived trustworthiness on whether to trust strangers (e.g., Chang, Doll, van't Wout, Frank, & Sanfey, 2010; Bente et al., 2012; Yu, Saleem, & Gonzalez, 2014; Li, Liu, Pan, & Zhou, 2017). Like the influence of perceived trustworthiness, physical attractiveness can also influence trust, with attractive people being perceived as more trustworthy (Yu et al., 2014). Thus, a host's perceived trustworthiness and attractiveness may influence decision making on Airbnb. Given the volume of studies relating perceived trust to price premiums in online markets (e.g., Ba & Pavlou, 2002; Houser & Wooders, 2006; Brown & Morgan, 2006; Gregg & Walczak, 2010) and that Ert et al. (2016) and (Jaeger et al., 2019) found partial support for the influence of perceived trustworthiness and attractiveness on prices in Airbnb, we hypothesize the following:

H1. Airbnb hosts perceived as trustworthy will have higher prices than Airbnb hosts perceived as untrustworthy.

H2. Airbnb hosts perceived as attractive will have higher prices than Airbnb hosts perceived as unattractive.

While Ert et al. (2016) and (Jaeger et al. (2019)) focused on the impact of attractiveness and trustworthiness on price, we draw on the theory of behavioural motivations (Gray, 1972) to develop hypotheses on potential interaction effects. Gray's theory on motivations characterizes the behavioural avoidance/inhibition system, which causes people to act cautiously when receiving cues of risk or danger until security is inferred (Gray, 1972; Park & Hinsz, 2006). Intuitively, proximity to and sharing a space with strangers poses a potential threat and results in greater anxiety and avoidance behaviour (Glick, DeMorest, & Hotze, 1988; Maeng & Tanner, 2013). Thus, the prospect of sharing a room or apartment with a host rather than renting an entire apartment is likely to activate the behavioural avoidance system. When the avoidance system is activated, people have vigilant processing (Watson, Wiese, Vaidya, & Tellegen, 1999), which may increase attention to facial cues related to the host's trustworthiness. Information processing also becomes more detailed as people feel more crowded by strangers (i.e., when sharing a room versus an apartment) (Maeng & Tanner, 2013). Thus, we hypothesize the following interaction between perceived trustworthiness and the listing type:

H3a. The perceived trustworthiness of hosts has a greater influence on prices for shared room listings compared to shared apartments and entire apartments.

H3b. The perceived trustworthiness of hosts has less influence on prices for entire apartment listings compared to shared rooms and shared apartments.

Park and Hinsz (2006) theoretically extend the model of behavioural motivation, which was developed based on the individual, to small

group environments. Park and Hinsz (2006) note that familiar groups (i.e., a group of people who know each other) are a source of security, i.e., there is "safety in numbers" (Scheidlinger, 1952), "herd instincts" are associated with safety motives, and that groups reduce the perceived risk of negative outcomes and fear (Baron, Kerr, & Miller, 1992). Thus, they propose that the perceived security of a group of familiar or known people decreases the activation of the behavioural avoidance system, lowering peoples' perceptions of threats, attention to threat cues, and vigilant information processing. The group size of a booking (i.e., the number of people allowed to rent a listing) varies on Airbnb. Given that an active avoidance system is likely to lead to greater focus on the perceived trustworthiness of a host and that the group size of the booking party is likely to reduce attention to threatening cues and motivation to process risk information, we hypothesize the following:

H4. The perceived trustworthiness of hosts has less influence on prices with larger group sizes of people occupying the rental.

3. Methodology

In this section, we present a research framework for integrating facial image analysis in tourism research by outlining six key steps in the research process (illustrated in Fig. 1). The research framework generally fits with other generalized processes for big data analytics (e.g. Gandomi; Haider, 2015). We examine each step, in turn.

3.1. Research problem

The first step in developing facial analytics research is outlining the research problem, including its motivation (both academic and practical), and articulating the research question. This should be theoretically and conceptually embedded through background literature, and clear definition and justification of research hypotheses to be empirically tested. This aspect was provided in the first two sections of the paper. The overarching research question is: What impact does the perceived trustworthiness and attractiveness of a host garnered from their profile photo have on the price of online accommodation, and how does this vary by accommodation type?

3.2. Identify and source data

Suitable sources of data to provide the empirical basis for testing the research hypotheses should be identified – high-quality data that will enable clear extraction of facial features using available algorithms. In some cases, this data may be provided by open access or via public sources or requested from researchers or organizations, in others, the data may be commercially available or potentially scraped from websites. If the images are not easily available, then it may be possible to identify images using scraping software from tools such as R, Python, and MATLAB. Once the image URLs are found, they can be downloaded using a bulk images downloader in most browsers.

For this study, data is available for many cities in a processed format with additional tools on the website insideairbnb.com. The data is available under a Creative Commons Universal (CC0 1.0) Public Domain Dedication license. We decided to focus on ten US cities that were popular on Airbnb: Boston, Chicago, Denver, Los Angeles, New Orleans, New York City, San Diego, San Francisco, Seattle, and Washington DC. The data covered city listings in the period from 3rd October to 19th October 2018. This included URLs for host images on 155,008 listings, which were then downloaded via MATLAB. A further 238 host images were found to be default (generic) user graphics. After excluding these images, 154,680 images remained.

3.3. Pre-process images

Data used in machine learning typically needs to be examined and, if



Fig. 1. Research framework: Steps in the recommended process for facial image analysis.

necessary, pre-processed. Traditional numeric data can be examined via visualisation and statistical analysis and may need treatment for missing values, outliers, standardization/normalization, or reclassification (see Baesens; Bapna, Marsden, Vanthienen, & Zhao, 2014). In the case of images, data that is not suitable for analysis needs to be screened out, and the images need to be formatted in a way that facilitates analysis, such as cropping the face. For faces, this procedure was facilitated by applying the Viola-Jones (Viola & Jones, 2001) algorithm, which systematically examines parts of the image with a cascade of binary classifiers for the face and sub-features and rejects it if facial features are not found (see Fig. 2). This includes properties common to human faces such as the eye region being darker than the upper cheeks, the nose bridge being lighter than the eyes, position, and size of facial features. By

applying a classification model that is based on upright and forward-facing facial features, such as Haar,¹ with a classification and regression tree (CART) analysis, we can screen out images that do not contain human faces, e.g. animals or buildings, or that are obscured, e.g. via lighting or various objects.

The Viola-Jones algorithm applies a variant Adaboost learning algorithm, which creates “strong” classifiers as a linear, weighted combination of “weak” classifiers (Viola & Jones, 2001). After trying various parameters, we set a MergeThreshold of 7 for face detection, which suppressed false detections by requiring target facial objects to be detected more times. Bounding boxes of detected faces are then cropped for computational efficiency. To improve accuracy, we also performed some manual screening for obvious false positives (i.e. classified as a host face but is not) and obvious false negatives (i.e. a host face that has been rejected). After processing using the Viola-Jones algorithm and some manual checking of the images in MATLAB, 96,554 faces remained.

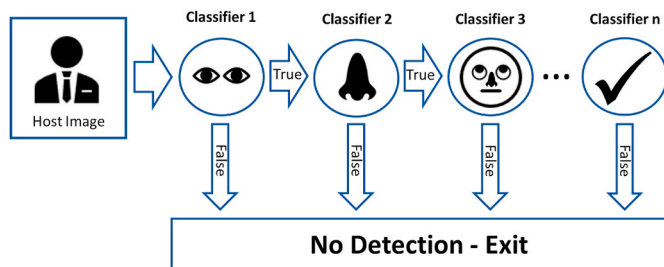


Fig. 2. Application of the Viola-Jones cascading classifier.

3.4. Deep learning model

Although training a Convolutional Neural Network (CNN) to identify types of facial images entirely from scratch is possible, numerous pre-trained models are available for facial classification, including network structures and pre-trained weights. Many advanced CNNs for image analytics are freely available in statistical software packages (e.g. GoogLeNet, ResNet-50, VGG19, and Inception-ResNet-v2) or via internet sources such as Github or the Caffe Model Zoo. CNNs are the

¹ Haar features are a sequential set of rescaled, box-shaped functions used in face recognition. Haar features are used for the detection of dark and light patches on a face to identify facial features. Haar features vary according to the facial characteristics being detected. For example, one feature involves a box with the bottom-half being dark and the top-half light to detect the eyes (dark) and forehead (light). Another involves a box with the left section being dark, the middle being light and the right section dark to detect the bridge of the nose (the light section in the middle). For more information about Haar features, please refer to Viola and Jones (2001).

most advanced and accurate methods for face recognition (Masi, Wu, Hassner, & Natarajan, 2018).

If the images for a piece of research are similar to those used for training the original network model being used, and the classification is identical, it can often be employed successfully 'straight out of the box' as a fixed feature extractor. If the images and/or classes are different, however, then the model will need to be retrained and/or retuned using transfer learning. Although we adopt a CNN designed for facial analysis, we consider the image types of Airbnb hosts to be slightly different than those that the VGG-Face CNN (Parkhi, Vedaldi, & Zisserman, 2015) were trained on (Huang, Mattar, Berg, & Learned-Miller, 2007; Wolf, Hassner, & Maoz, 2011), and therefore we fine-tune the VGG-Face CNN in this study. The CNN adopted in this study, VGG-Face, was developed by the Visual Geometry Group at the University of Oxford specifically for face classification tasks. It was found to have superior accuracy to other face CNNs, estimated at 97.3% when tested on the YouTube Faces Dataset (Parkhi et al., 2015).

To retrain the VGG-Face CNN for two separate networks classifying trustworthy/untrustworthy and attractive/unattractive faces, we used the MIT Faces dataset developed by Bainbridge, Isola, and Oliva (2013). The database contains a diverse set of images of 2222 persons according to gender, age, and ethnicity that are broadly representative of adult US citizens. Within the dataset, all 2222 faces were rated for attractiveness on a scale of 1 = unattractive to 5 = attractive. Further, the faces were rated for various psychological attributes, including trustworthiness, on a scale of 1 = not at all to 9 = extremely. For the purposes of retraining our classifier, we used faces classified at each end of the rating scales, excluding ambiguous faces. Since no faces were rated 8 or 9 on the trustworthiness scale, we used those rated 7 to 9. Similarly, we included untrustworthy faces rated 1 to 3. Very few faces were rated 1 or 5 for attractiveness, and thus we used faces rated 1 to 2 and 4 to 5 for retraining the CNN.

Images were automatically resized for the VGG-Face image input layer to a size of 224-by-224-by-3 using an augmented datastore, where 3 is the number of colour channels. Convolutional CNN layers extract features of images that the final learnable layer and classification layer apply for image classification. Early layers learn low-level features, such as blobs, colours, and edges, while the final layers learn very specific details for tasks. To retrain the network to learn features specific to the MIT faces dataset so that we can classify new Airbnb host images, the last two layers were replaced with those adapted for the new face data. To learn faster in the new learnable layer than in the transferred layers, we use an increased learning rate factor of the learnable layer (new_fc in Fig. 3). The new classification layer specifies the output classes of the network during training (new_classoutput in Fig. 3). We froze weights of the first ten layers in the network by setting their learning rates to zero, both to speed-up training and to avoid overfitting to the new data. Moreover, augmentation procedures on training images helped the network to circumvent memorizing precise details of training images and avoid overfitting (randomly translate images by up to 30 pixels, randomly flip images vertically, and rescale images vertically and horizontally by up to 10%).

The network used is shown in Fig. 3. Our convolutional layers apply sliding convolutional filters to inputs to detect image features. Examples of image features extracted from convolutional layers for the trustworthiness network are given in Fig. 4. Rectified Linear Unit (ReLU) layers apply a threshold operation to each element of input, such that $f(x) = \{x \text{ if } x \geq 0 \text{ and } 0 \text{ if } x < 0\}$. Max pooling divides an input into regions and calculates the maximum for each (down-sampling). Fully connected layers multiply the input by a weight matrix and add a bias vector. Our dropout layers randomly set input elements to zero with a 50% probability. The softmax layer applies a softmax function (multiclass generalized logistic sigmoid function; Bishop, 2006). The classification layer uses values from the softmax function to assign each host image to one of the two classes using the cross-entropy function. In terms of training options for the CNN, we set the learning rate to 0.0003 to slow down

learning in the non-frozen transferred layers. This enables fast learning in the new layers, slower learning in the middle layers, and no learning in the top frozen layers. We set the number of training epochs to 6, with 78 iterations per epoch (epochs are full training cycles on the entire dataset; fewer epochs are needed for transfer learning).

We then trained the network using the training data, consisting of 70% of the MIT face data set described above. Subsequently, we classified the validation images using the fine-tuned network (30% of the data set was used for validation), calculated the classification accuracy, and examined sample image output for face validity. All sample images appeared to be accurately classified. The accuracy for the retrained CNN for trustworthiness was 99.17%, while for attractiveness the accuracy was 91.07%. It took nearly 4 h to retrain the network on a very high-specification MacBook Pro computer. Examples of features extracted from convolutional layers of the trustworthiness network are given in Fig. 4.

The trained CNNs were used to classify the 96,554 host faces. A total of 76,007 (78.72%) of faces were classified as trustworthy, while 20,547 (21.28% were classified as untrustworthy). A much smaller number, 39,651 (41.07%) of faces were classified as attractive, with 56,903 (59.93%) classified as unattractive.

3.5. Assemble data

Once the deep learning data has been created, it needs to be combined with other research data before the analysis can begin. If the traditional data set includes the URL or name of the image, then the name of the image can be used as a key for joining the facial emotion data set with the traditional data (e.g., host profile, accommodation characteristics, and review scores) in statistical packages such as SPSS, MATLAB or R. In the case of our example, the Airbnb data from InsideAirbnb include the host image URL from which the image name was extracted to enable merging the data sets. However, since some of the image names were identical, this caused a problem; as a solution, we added a unique numerical identifier to each image name for key matching. Further to joining the data sets, a further round of reformatting and cleaning may be needed, depending on the intended type of statistical analysis.

3.6. Analyse data and test hypotheses

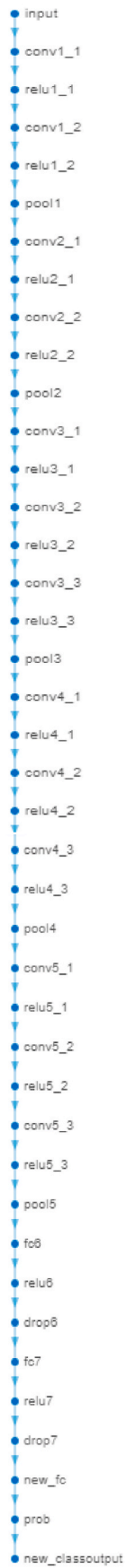
Many forms of analysis can potentially be used to test the research hypotheses for facial image studies, depending on the nature of the hypotheses and the data being used. For example, testing for differences might employ ANOVA, while examining the significance of statistical relationships for multiple variables could use a suitable form of regression or structural equation modelling. In this study, we use linear regression to test the research hypotheses, as reported in the next section.

4. Results

In this section, we use our dataset to test the hypothesized effects of trustworthiness and attractiveness on housing prices developed in Section 2. We also provide analysis on features that impact pricing and compare our results on trustworthiness and attractiveness to the studies of Ert et al. (2016) and Wang and Nicolau (2017).

4.1. Data and descriptive statistics

Primarily, we are interested in using our methodology to determine the impact of perceived trustworthiness and attractiveness on rental prices. As many factors may influence the price of Airbnb listings, Wang and Nicolau (2017) study twenty-five explanatory variables in five categories – host attributes, site and property attributes, amenities and services, rental rules, and online review ratings – across 33 cities to



	NAME	TYPE	ACTIVATIONS	LEARNABLES
1	input 224x224x3 images	Image Input	224x224x3	-
2	conv1_1 64 3x3x3 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	224x224x64	Weights 3x3x3x64 Bias 1x1x64
3	relu1_1 ReLU	ReLU	224x224x64	-
4	conv1_2 64 3x3x64 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	224x224x64	Weights 3x3x64x64 Bias 1x1x64
5	relu1_2 ReLU	ReLU	224x224x64	-
6	pool1 2x2 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	112x112x64	-
7	conv2_1 128 3x3x64 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	112x112x128	Weights 3x3x64x128 Bias 1x1x128
8	relu2_1 ReLU	ReLU	112x112x128	-
9	conv2_2 128 3x3x128 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	112x112x128	Weights 3x3x128x128 Bias 1x1x128
10	relu2_2 ReLU	ReLU	112x112x128	-
11	pool2 2x2 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	56x56x128	-
12	conv3_1 256 3x3x128 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	56x56x256	Weights 3x3x128x256 Bias 1x1x256
13	relu3_1 ReLU	ReLU	56x56x256	-
14	conv3_2 256 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	56x56x256	Weights 3x3x256x256 Bias 1x1x256
15	relu3_2 ReLU	ReLU	56x56x256	-
16	conv3_3 256 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	56x56x256	Weights 3x3x256x256 Bias 1x1x256
17	relu3_3 ReLU	ReLU	56x56x256	-
18	pool3 2x2 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	28x28x256	-
19	conv4_1 512 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	28x28x512	Weights 3x3x256x512 Bias 1x1x512
20	relu4_1 ReLU	ReLU	28x28x512	-
21	conv4_2 512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	28x28x512	Weights 3x3x512x512 Bias 1x1x512
22	relu4_2 ReLU	ReLU	28x28x512	-
23	conv4_3 512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	28x28x512	Weights 3x3x512x512 Bias 1x1x512
24	relu4_3 ReLU	ReLU	28x28x512	-
25	pool4 2x2 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	14x14x512	-
26	conv5_1 512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	14x14x512	Weights 3x3x512x512 Bias 1x1x512
27	relu5_1 ReLU	ReLU	14x14x512	-
28	conv5_2 512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	14x14x512	Weights 3x3x512x512 Bias 1x1x512
29	relu5_2 ReLU	ReLU	14x14x512	-
30	conv5_3 512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	14x14x512	Weights 3x3x512x512 Bias 1x1x512
31	relu5_3 ReLU	ReLU	14x14x512	-
32	pool5 2x2 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	7x7x512	-
33	fc6 4096 fully connected layer	Fully Connected	1x1x4096	Weights 4096x25088 Bias 4096x1
34	relu6 ReLU	ReLU	1x1x4096	-
35	drop6 50% dropout	Dropout	1x1x4096	-
36	fc7 4096 fully connected layer	Fully Connected	1x1x4096	Weights 4096x4096 Bias 4096x1
37	relu7 ReLU	ReLU	1x1x4096	-
38	drop7 50% dropout	Dropout	1x1x4096	-
39	new_fc 2 fully connected layer	Fully Connected	1x1x2	Weights 2x4096 Bias 2x1
40	prob softmax	Softmax	1x1x2	-
41	new_classoutput crossentropyx with classes 'Trustworthy' and 'Untrustworthy'	Classification Output	-	-

Fig. 3. Network Configuration.

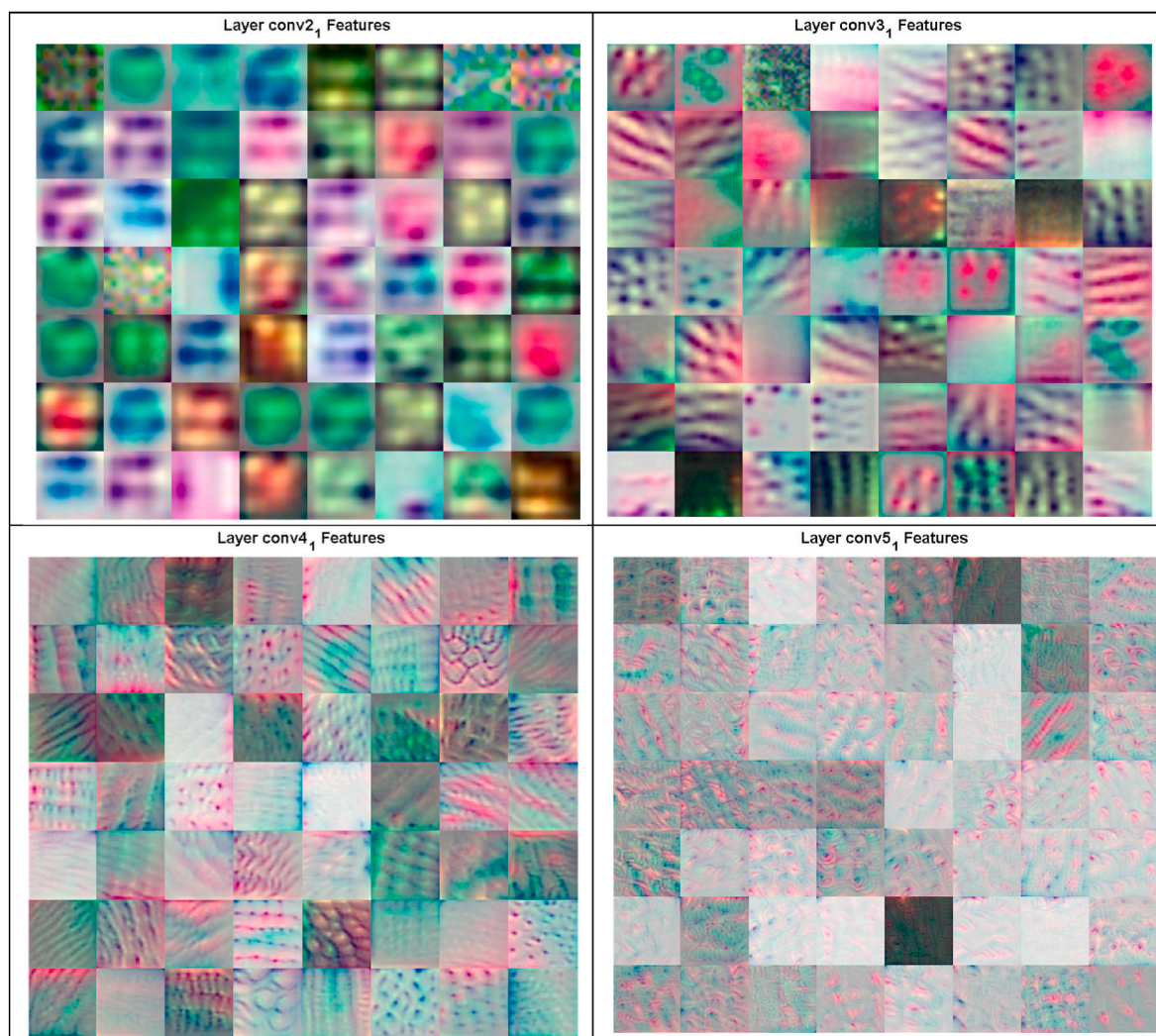


Fig. 4. Example features from trained trustworthiness CNN.

determine the impact of these factors on price. The twenty-five variables selected correspond to the information provided by the third-party website, [insideairbnb.com](https://www.insideairbnb.com). As we also sourced our data from the same third-party website, we consider the same factors utilized in Wang and Nicolau (2017) as control variables in our study. Table 1 presents the relevant variable names, definitions, and descriptive statistics (mean and standard deviation) of the data collected. In Section 4.5, we compare the impact of the control variables on prices with the results of Wang and Nicolau (2017).

From our total dataset of 96,554 observation, we removed any observations that had missing data from either ratings, price, the number of bathrooms, and the number of bedrooms (and we one observation that corresponded to a location in Houston). This left us with a dataset of 78,215 observations, where 79.32% of hosts were classified as trustworthy, while 40.41% were classified as attractive.

4.2. Hypotheses testing

To test our hypotheses, we consider the following four regression models: Model 1 has perceived trustworthiness as an independent variable; Model 2 has perceived attractiveness as an independent variable; Model 3 has both trustworthiness and attractiveness as independent variables; Model 4 uses both variables and adds an interaction term. Note that these four models are the same that are presented in the analysis of Section 4 of Ert et al. (2016). For each model, the dependent

variable is the log of daily rental prices, as the distribution of rental prices is heavily skewed. We consider a comprehensive set of control variables, which are consistent with the variables of Wang and Nicolau (2017). Table 2 shows the results of the four models.

Trustworthiness and attractiveness are significant either when one of the variables (Model 1: trust only, $\beta = 0.0126$, $p < 0.001$; attractiveness only, $\beta = 0.0085$, $p < 0.001$) or both variables are included in the model (Model 3: trust, $\beta = 0.0106$, $p < 0.001$; attractiveness, $\beta = 0.0061$, $p < 0.001$), and there is no evidence of an interaction between the two factors (Model 4: trust x attractiveness, $\beta = -0.0031$, $p > 0.05$). Across the four models, using the average values from Table 1 and varying the presence of trustworthiness and attractiveness, the effect size for perceived trustworthiness ranges from a 2.47%–2.94% increase in price, while for attractiveness, it ranges from a price increase of 1.41%–2.05%. For both facial characteristics, the price increase ranged from 3.92% to 4.71%. Thus, people perceived as trustworthy and attractive in our dataset, *ceteris paribus*, can obtain a higher price by approximately 5% compared to those perceived as untrustworthy and unattractive.

All four models demonstrate that perceived trustworthiness and perceived attractiveness impact prices, providing support for H1 and H2. This simultaneously contrasts and supports the results from Ert et al. (2016) and (Jaeger et al. (2019)), as Ert et al. (2016) found support for trustworthiness and (Jaeger et al. (2019)) found support for attractiveness. Whereas Ert et al. (2016) found minor support for an interaction effect between attractiveness and trustworthiness ($p \approx 0.09$), we find no

Table 1
Variables used in the research.

Variable Name	Mean	Std. Dev.	Definition
Trust	0.7932	0.4050	If the host is classified as trustworthy (DV)
Attract	0.4041	0.4907	If the host is classified as attractive (DV)
Price	154.20	230.80	Listed price per night (U.S. Dollars)
Super Host	0.3250	0.4684	Host has a super host badge (DV)
Host No. Listing	7.2284	37.955	Host's number of accommodation rentals listed in Airbnb.com
Host Verified	0.5589	0.4965	The host has ID verified by Airbnb (DV)
Review rate	1.9462	2.0438	Average of reviews per month since rental was first listed
(Reviews Rate) ²	7.9650	19.517	Review rate squared
Review Scores	94.924	7.4522	Overall reviews scores
No. Reviews	36.574	54.745	Total number of reviews
Instant Book	0.4152	0.4928	Offers instant bookable (DV)
Accommodates	3.4877	2.3628	The number of people that can be accommodated
No. Bedrooms	1.3647	0.9633	The number of bedrooms
No. Bathrooms	1.2910	0.6832	The number of bathrooms
Real Bed	0.9828	0.1298	Offers a real bed (versus couches, airbeds, etc.) (DV)
House	0.2540	0.4353	Accommodation is bungalow, house, townhouse, villa, cabin and chalet (DV)
Apartment	0.6215	0.4850	Accommodation is Apartment, condominium and loft (DV)
Other properties	0.1245	0.3302	Other types of accommodation (e.g., B&B, dorm) (Reference group)
Entire Home	0.6157	0.4864	Entire accommodation for guest (DV)
Private Room	0.3610	0.4803	Shared accommodation with host (DV)
Shared room	0.0232	0.1506	Guest shares room with host or other guest (Reference group)
Cancellation Policy	0.7696	0.4211	No cancellation or penalty applies. (DV)
Free Wi-Fi	0.9827	0.1303	Offers wireless internet (DV)
Free Parking	0.3562	0.4789	Offers free parking (DV)
Breakfast	0.1118	0.3152	Offers breakfast (DV)
Smoking Allowed	0.0505	0.2189	Smoking is allowed (DV)
Picture verification	0.0302	0.1712	Require guest's profile picture for booking approval (DV)
Phone verification	0.0379	0.1910	Require guest's phone number for booking approval (DV)
Distance	9.1698	8.2936	Distance (Km) between the listing and city centre

Note. (DV) indicates dummy variable. For consistency with the literature, the variable No. Reviews is only used in the analysis related to variables from [Ert et al. \(2016\)](#) (See 4.4).

support for any interaction ($\beta = -0.0001$, $p > 0.05$), likely because attractiveness is independently significant in our dataset.

Next, we explore interaction effects between perceived trustworthiness and features of the listing. [Table 3](#) shows the regression results for three models based on trust. For brevity, we present only the variables related to H3 and H4 (the remaining coefficients and standard errors are similar to [Table 2](#) when adding the interaction effects). Model 1 adds interaction effects between trust and renting a private room and trust and having an entire home to Model 3 from [Table 2](#) (i.e., the control variables from [Wang and Nicolau \(2017\)](#) with both trustworthiness and attractiveness, but not the interaction between them). Both interaction effects are significant and negative (Model 1: Trust x Entire Home, $\beta = -0.0314$, $p < 0.01$; Trust x Private Room, $\beta = -0.0215$, $p < 0.05$), implying that the impact of trust is greatest for listings that involve sharing a room with a stranger. Furthermore, the results show that there is a greater reduction on the impact of perceived trustworthiness for entire homes compared to private rooms. Model 2 considers the interaction between trust and the number of bedrooms. Again, the interaction effect between trust and the No. of Bedrooms is negative and significant (Trust x No. Bedrooms, $\beta = -0.0064$, $p < 0.001$), implying that the more bedrooms there are, the less impact perceived trust has on price. Based on the effect sizes, a listing with three bedrooms eliminates

the influence of perceived trustworthiness. Finally, Model 3 considers all three interactions. The results show that in the combined model, all three interaction effects remain negative and significant (Trust x Entire Home, $\beta = -0.0287$, $p < 0.01$; Trust x Private Room, $\beta = -0.0213$, $p < 0.05$, Trust x No. Bedrooms, $\beta = -0.0051$, $p < 0.01$). Together, these results provide support for H3a, H3b, and H4.

4.3. Relationship between review rates and price

In this subsection, we explore the potential for a quadratic relationship between review rates and price. [Wang and Nicolau \(2017\)](#) found that the number of reviews per month was negatively correlated with price, reasoning that this results from cheaper listings receiving more bookings and reviews. This negative relationship between the number of reviews and price is supported by [Benítez-Aurioles's \(2018\)](#) study of 44 cities across the world, by [Gibbs, Guttentag, Gretzel, Morton, and Goodwill's \(2018\)](#) study of 5 cities in Canada, and [Chattopadhyay and Mitra's \(2019\)](#) study of 11 American cities. While we find that the review rate is negatively related to price, we also investigate the potential for a non-linear relationship.

Our models include a quadratic relationship with the price since we posit that prices may initially decrease with increases in the rate of reviews, but then start increasing with greater review rates. Online reputation and rating systems often lead a "the rich get richer" phenomenon ([Hu, Milner, & Wu, 2015](#)). For example, [Carare \(2012\)](#) found that consumers' willingness to pay for top-ranked mobile applications on Apple's App Store was greater when the ranking was provided, but that the bestseller status effect (which held for the top-50 apps) on willingness to pay declined steeply with rank. Thus, as a threshold of popularity is reached, the most popular properties may be able to charge more due to increased demand pressure, which is in line with the "tipping point" theory of social behaviour ([Gladwell, 2000](#)).

Furthermore, rental accommodations such as Airbnb units and hotel rooms are perishable resources. Thus, similar to hotels, hosts have a fixed capacity and, as in classic revenue management problems (e.g., hotels and airlines), a host's primary mechanism for reacting to an increase (or decrease) in customer demand is by raising (or lowering) price ([Talluri & Van Ryzin, 2006](#)). Although hosts do not have access to the sophisticated platforms to make optimal decisions, experimental evidence suggests that non-experts (e.g., students in lab experiments) can perform well when constrained by limited capacity in revenue management and pricing tasks ([Bearden, Murphy, & Rapoport, 2008](#); [Bendoly, 2013](#); [Kocabiyikoglu, Gogus, & Gonul, 2015](#); [Kremer, Mantin, & Ovchinnikov, 2017](#)). For example, in a finite horizon revenue management task with limited capacity, [Bearden et al. \(2008\)](#) found that "decision-makers employ policies of the same form of the optimal policy." More recently, [Kremer et al. \(2017\)](#) found in a pricing task that after gaining experience, people priced optimally facing myopic consumers and performed well given a mix of myopic and strategic consumers. As hosts have ample opportunities to learn, we posit that hosts will have lower prices initially at high ratings, but at a point, higher ratings will correspond to an increase in prices.

[Table 2](#) shows that a higher review rate corresponds to a lower price (Model 3: Review Rate, $\beta = -0.0210$, $p < 0.001$). (Given the non-significant interaction term, we suggest referring to Model 3). However, we also added the square of the review rate to capture potential quadratic effects in this analysis. [Table 2](#) shows that there is a significant and positive relationship between prices and the square of the review rate (Model 3: (Review Rate)², $\beta = 0.0007$, $p < 0.001$), providing support for the quadratic relationship. Thus, hosts may respond to high bookings levels by increasing prices.

4.4. Comparison to [Ert et al. \(2016\)](#)

[Ert et al. \(2016\)](#) considered a reduced set of control variables compared to our models. Thus, for robustness, we re-ran the four models

Table 2
Regression results using data variables from Wang and Nicolau (2017).

	Model 1		Model 2		Model 3		Model 4	
(Intercept)	1.1152***	[0.0126]	1.121***	[0.0125]	1.1149***	[0.0126]	1.1146***	[0.0126]
Trust	0.0126***	[0.0017]			0.0106***	[0.0018]	0.0112***	[0.0020]
Attract			0.0085***	[0.0014]	0.0061***	[0.0015]	0.0088*	[0.0042]
Trust*Attract							-0.0031	[0.0045]
Super Host	0.0177***	[0.0017]	0.0183***	[0.0017]	0.0180***	[0.0017]	0.0180***	[0.0017]
Host No. Listing	0.0004***	[<0.0001]	0.0004***	[<0.0001]	0.0004***	[<0.0001]	0.0004***	[<0.0001]
Host Verified	0.0058***	[0.0014]	0.0060***	[0.0014]	0.0058***	[0.0014]	0.0058***	[0.0014]
Reviews Rate	-0.0211***	[0.0006]	-0.0211***	[0.0006]	-0.0210***	[0.0006]	-0.0210***	[0.0006]
(Review Rate) ²	0.0007***	[0.0001]	0.0007***	[<0.0001]	0.0007***	[<0.0001]	0.0007***	[<0.0001]
Review Scores	0.0026***	[0.0001]	0.0026***	[0.0001]	0.0026***	[0.0001]	0.0026***	[0.0001]
Instant Book	-0.0053***	[0.0015]	-0.0056***	[0.0015]	-0.0054***	[0.0015]	-0.0054***	[0.0015]
Accommodates	0.0305***	[0.0005]	0.0305***	[0.0005]	0.0305***	[0.0005]	0.0305***	[0.0005]
No. Bedrooms	0.0606***	[0.0012]	0.0606***	[0.0012]	0.0606***	[0.0012]	0.0606***	[0.0012]
No. Bathrooms	0.0501***	[0.0013]	0.0501***	[0.0013]	0.0501***	[0.0013]	0.0501***	[0.0013]
Real Bed	0.0112*	[0.0054]	0.0110*	[0.0054]	0.0111*	[0.0054]	0.0111*	[0.0054]
House	-0.0171***	[0.0024]	-0.0171***	[0.0024]	-0.0171***	[0.0024]	-0.0171***	[0.0024]
Apartment	0.0021	[0.0023]	0.0014	[0.0023]	0.0017	[0.0023]	0.0017	[0.0023]
Entire Home	0.5118***	[0.0049]	0.5118***	[0.0049]	0.5115***	[0.0049]	0.5115***	[0.0049]
Private Room	0.2671***	[0.0048]	0.2670***	[0.0048]	0.2669***	[0.0048]	0.2669***	[0.0048]
Cancellation Policy	0.0153***	[0.0017]	0.0153***	[0.0017]	0.0152***	[0.0017]	0.0152***	[0.0017]
Free Wi-Fi	0.0073	[0.0053]	0.0077	[0.0053]	0.0074	[0.0053]	0.0074	[0.0053]
Free Parking	0.0044*	[0.0017]	0.0043*	[0.0017]	0.0043*	[0.0017]	0.0043*	[0.0017]
Breakfast	0.0220***	[0.0022]	0.0219***	[0.0022]	0.0219***	[0.0022]	0.0219***	[0.0022]
Smoking Allowed	-0.0148***	[0.0032]	-0.0155***	[0.0032]	-0.0148***	[0.0032]	-0.0148***	[0.0032]
Picture Verification	-0.0169*	[0.0069]	-0.0175*	[0.0069]	-0.0172*	[0.0069]	-0.0172*	[0.0069]
Phone Verification	0.0282***	[0.0062]	0.0283***	[0.0062]	0.0283***	[0.0062]	0.0283***	[0.0062]
Distance	-0.0027***	[0.0001]	-0.0026***	[0.0001]	-0.0027***	[0.0001]	-0.0027***	[0.0001]
R ² (Adj)	0.604		0.604		0.604		0.604	
N	78,215		78,215		78,215		78,215	

Note: **p* < 0.05, ***p* < 0.01, ****p* < 0.001. Standard Errors in square brackets. City variables used as controls. Shared Room is reference group.

Table 3
Regression results for trust interaction effects.

	Model 1		Model 2		Model 3	
Trust	0.0190***	[0.0030]	0.0372***	[0.0103]	0.0423***	[0.0105]
Trust*Entire Home	-0.0314**	[0.0105]			-0.0287**	[0.0106]
Trust*Private Room	-0.0215*	[0.0106]			-0.0213*	[0.0106]
Trust*No. Bedrooms			-0.0064***	[0.0019]	-0.0051**	[0.0019]
R ² (Adj)	0.604		0.604		0.604	
N	78,215		78,215		78,215	

Note: **p* < 0.05, ***p* < 0.01, ****p* < 0.001. Standard Errors in square brackets. With all control variables. Shared Room is reference group.

above using the control variables considered by Ert et al. (2016). The results, which are presented in Table 4, are consistent with the previous analysis (which contain the control variables from Wang & Nicolau, 2017). Thus, the results from both sets of models (i.e., Tables 4 and 2) provide support for both H1 and H2: trustworthy and attractive photos are associated with higher prices.

To demonstrate the overall consistency of our dataset with Ert et al.

(2016), we also analyse the impact of the control variables on the listing price. Our results largely resemble that of Ert et al. (2016). We find that entire homes lead to a significant increase in price relative to shared and private rooms, private rooms have higher prices than shared rooms (Ert et al., 2016 did not find a significant relationship here), and larger apartments (measured by the number of bedrooms) also correspond to higher prices (Model 4: entire home, $\beta = -0.5202$, $p < 0.001$; private

Table 4
Regression results using data variables from Ert et al. (2016).

	Model 1		Model 2		Model 3		Model 4	
Intercept	1.2949***	[0.0107]	1.2473***	[0.0106]	1.2407***	[0.0107]	1.2407***	[0.0107]
Trust	0.0156***	[0.0018]			0.0121***	[0.0019]	0.0121***	[0.0019]
Attract			0.0133***	[0.0015]	0.0107***	[0.0015]	0.0108**	[0.0045]
Trust*Attract							-0.0001	[0.0047]
Review Ratings	0.0022***	[0.0001]	0.0022***	[0.0001]	0.0022***	[0.0001]	0.0022***	[0.0001]
No. Bedrooms	0.1322***	[0.0008]	0.1325***	[0.0008]	0.1323***	[0.0008]	0.1323***	[0.0008]
Entire Home	0.5202***	[0.0049]	0.5200***	[0.0049]	0.5196***	[0.0049]	0.5196***	[0.0049]
Private Room	0.2229***	[0.0049]	0.2230***	[0.0049]	0.2227***	[0.0049]	0.2227***	[0.0049]
No. Reviews	-0.0002***	[<0.0001]	-0.0002***	[<0.0001]	-0.0002***	[<0.0001]	-0.0002***	[<0.0001]
R ² (Adj)	0.559		0.559		0.559		0.559	
N	78,215		78,215		78,215		78,215	

Notes: **p* < 0.05, ***p* < 0.01, ****p* < 0.001. Standard Errors in square brackets. City variables used as controls. Shared Room is reference group.

room, $\beta = -0.2229$, $p < 0.001$).

Ert et al. (2016) also found that neither review scores nor the volume of reviews impact prices. However, our results consistently show that higher review scores increase the price ($\beta = 0.0022$, $p < .001$), while a higher number of reviews corresponds to lower prices ($\beta = -0.0002$, $p < .001$). The latter is consistent with the reputation effect from our earlier analysis as well as other articles in the literature (e.g., Benítez-Aurioles, 2018; Chattopadhyay & Mitra, 2019; Gibbs, Guttentag, Gretzel, Morton, & Goodwill, 2018; Wang & Nicolau, 2017) that look at the monthly (or annual) rate of reviews.

4.5. Comparison to Wang and Nicolau (2017)

Finally, we compare our results to the findings of Wang and Nicolau (2017). As previously mentioned, our dataset captured the set of variables considered in the analysis by Wang and Nicolau (2017) (We note that Wang and Nicolau (2017) also include a variable for the host having a photo; however, each host in our dataset has a profile photo.). Table 2 can be used for comparing our results to Wang and Nicolau (2017). Overall, the direction and significance of most variables in our analysis align with the values of Wang and Nicolau (2017), as shown in Table 2 (again, given the non-significant interaction term, we suggest referring to Model 3). In terms of host characteristics, prices are higher for a Super Host, the host's identity being verified, and a host having a greater number of listings. Regarding the accommodation, being closer to the city centre, offering an entire home or a private room, offering real beds, having more bedrooms and more bathrooms, thereby accommodating more people, leads to higher prices. For amenities and policies, free parking, guests requiring phone verification, no smoking, and strict cancellation policies also correspond to higher prices. The only places where our results differ from Wang and Nicolau (2017) is regarding breakfast (they find offering breakfast decreases price, where we find it increases the price), Wi-Fi (we do not find it significant), and we find that apartments have higher prices compared to houses (which could be due to differences in the cities selected for the studies). Thus, our analysis largely supports the results of Wang and Nicolau (2017).

5. Conclusions

5.1. Summary of findings and contributions

Using a large set of host images classified using a retuned facial convolutional neural network, our research shows that both host trustworthiness and attractiveness command a price premium for Airbnb accommodation. We go well beyond previous inconclusive studies to provide a comprehensive test of both effects. By examining interaction effects, we provide new insight into the scope of the impact of trust on prices for types of accommodation. More explicitly, we extend previous work by testing Gray's theory of motivation via the examination of pricing for different types of accommodation, finding that trust is more important in situations of smaller accommodations and that shared with strangers.

Our paper has three main contributions. First, we make an original contribution to knowledge by testing for interactions between trustworthiness, attractiveness, and other control variables via Gray's (1972) motivation theory. This is an original application of Gray's theory in tourism research to examine whether sharing a home with a host (or other strangers) increases the effects of trustworthiness on price and whether larger accommodations (measured by the number of bedrooms) dampen the effect of trust on price. Notably, previous studies have not examined the interaction of perceived trust and perceived attractiveness based on facial features according to the type of accommodation.

Our second major contribution is methodological. We develop a simple, original framework to enable advanced facial analytics for academic inquiries by tourism researchers. While text analytics is widely used in the social sciences, image analytics is not widespread.

Methodologies for image analytics are still an active area of research, and using developed methods often requires a high level of specialized knowledge. In contrast, our framework provides a guide for integrating facial image analytics in tourism research. Rather than using MTurk workers or students, we utilize deep learning (via a convolutional neural network) to classify host profile photos as attractive (or unattractive) and trustworthy (or untrustworthy). Consequently, our approach offers two benefits over existing methods. First, once "exemplar" human ratings of images have been created they can be used for training a Convolutional Neural Network (CNN) and then used consistently for assessment of images, rather than generating human-rated images ad hoc for every study. Thus, once a CNN has been trained, it can be reapplied in the same context at near-zero cost, providing cost benefits over ad hoc survey methods. Second, using artificial intelligence methods and computational power, we can process a much larger sample of images. For our study, which is focused on the US, we use a substantial dataset of 78,215 Airbnb listings across 10 American cities and train our CNN using a data set of American faces.

Our third major contribution is to provide practical insight into the effect of the perceived trustworthiness and attractiveness of host images based on facial features on the prices that can be charged by Airbnb providers for accommodation. Our findings equip hosts with new knowledge to improve how their profile is conveyed through the online accommodation platform to obtain higher prices. Further, new insights into interaction effects according to the types of accommodation, provide a unique understanding of the importance of perceived trustworthiness based on facial features in achieving higher prices for smaller accommodation and accommodation that is shared with strangers.

Finally, as an additional original insight, we find support for a proposed quadratic relationship between the number of reviews and pricing, suggesting a "tipping point" of review popularity that subsequently impacts pricing. Prices initially decrease with increases in the rate of reviews, but then start increasing with greater review rates, presumably as a threshold of popularity is reached whereby the most popular properties can charge more due to increased pressure on demand.

5.2. Implications for practice

The results underline the importance of well-chosen, pre-screened host images on accommodation websites. Appropriate host images can command a price premium (around 5% in our research), increasing revenues. This selection can be improved by the application of trained CNNs that recognize appropriate characteristics. Such a tool could be built into platforms and applied at the accommodation profile building stage.

Further, the impact of host images that convey trust varies by property type and is particularly important for small accommodation and where it is shared with strangers. In these situations, sensitivity analysis could reveal the level of the price premium that can be commanded by a host/property. The level of the price premium can be automatically calculated and built into the pricing mechanism of accommodation platforms.

Sharing economy platforms such as Uber have built-in algorithms to vary price according to demand, referred to as surge pricing. Our research has suggested the need to build similar algorithmic mechanisms into accommodation platforms to translate increased demand, captured by high volumes of reviews, into increased prices in a more systematic way. This represents yet another aspect that accommodation platforms can improve using data science to increase revenues and profits.

5.3. Limitations of the study

There are several limitations to the study. The first of these relates to the data that we used to test the hypotheses. While the dataset can be considered very large by traditional tourism research standards, it is

based on ten US cities. Since the CNN is also trained on US data, this adds validity to classification in the research. However, other (excluded) US cities may vary in their accommodation characteristics and host populations, and therefore it would be desirable to test the hypotheses with data from a larger set of US cities. In addition, our research is limited to the US context. Training and testing the network in other countries may find different results but would also require training data and test images from the specific country context.

Second, our results suggest that the network for trust is more accurate than that of attractiveness. This could be due to limitations of the training dataset. Although broadly representative of the US population, the training dataset is limited in terms of its size and the availability of extreme pole data. The face data rated at each end of the semantic scales were either small or non-existent, requiring a broader set of rating points for developing the training datasets. The results may also be due to differences in perceptions of attractiveness among raters than of trust. Again, more varied data may improve accuracy. In addition, another facial CNN could provide more accurate classification.

Third, our analysis restricts classification to be binary (e.g., trustworthy or untrustworthy). Greater insights in the role of trustworthiness and attractiveness on prices could be realized with greater granularity in the classification of these features (i.e., having scales from 1 to 5 or 1 to 9, rather than binary outcomes). However, to achieve this granularity would require a substantially larger training dataset with a greater variety of trustworthiness and attractiveness. Recall that the CNN was trained on 2222 faces. Moreover, no faces rated above 7 (out of 9) for trustworthiness, and few faces rated as 1 or 5 (out of 5) for attractiveness. Future research could create or leverage an alternative and larger-scale training dataset to classify faces on a more detailed scale to determine the additional value of larger data for setting listing prices.

Fourth, our research focuses on the role of trustworthiness and attractiveness, which is aligned with a significant body of research studying inferences from photos and appearance. However, research shows that social evaluation can consist of two dimensions: valence/trustworthiness and power/dominance (Todorov, Said, Engell, & Oosterhof, 2008; Stewart et al., 2012). Given Gray's theory of motivation theory, cues of power/dominance could likely result in behavioural avoidance, which could influence the host's prices, particularly for smaller or shared accommodations. Thus, future research could extend our research to look at the impact of power/dominance and its potential interaction with valence/trustworthiness on prices in Airbnb.

Fifth, our research is entirely focused on the perceived trustworthiness and attractiveness of hosts based on photos. An interesting direction for future work would be to develop mechanisms that can capture true trustworthiness and attractiveness, incorporating these factors into platforms like Airbnb to determine their impact on pricing. Finally, the CNN was retrained using a single computer. The use of high-performance computing facilities, e.g. through access to cloud-based parallel processing, would reduce training time considerably.

5.4. Concluding comments

This paper introduces an original framework for applying deep learning approaches for facial image analytics using big data to tourism research. With the inclusion of deep learning packages in many types of modern statistical software, deep learning analytics is now an accessible means of obtaining theoretical and practical insights into tourism phenomena. We hope that the simplicity of the framework encourages other tourism researchers to adopt a similar approach in their future work.

Impact statement

Our research provides practical insight into the effect of the perceived trustworthiness and attractiveness of host images based on facial features on the prices that can be charged by Airbnb providers for accommodation. Our findings equip hosts with new knowledge to

improve how their profile is conveyed through the online accommodation platform to obtain higher prices. Further, new insights into interaction effects according to the types of accommodation, provide a unique understanding of the importance of perceived trustworthiness based on facial features in achieving higher prices for smaller accommodation and accommodation that is shared with strangers.

Credit author statement

Stuart J. Barnes: Conceptualization; Software; Methodology; Visualisation; Validation; Formal analysis; Resources; Project administration; Investigation; Writing – original draft; Writing – review and editing. Samuel Nathan Kirshner, Conceptualization; Software; Methodology; Visualisation; Validation; Formal analysis; Resources; Project administration; Investigation; Writing – original draft; Writing – review and editing.

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.tourman.2020.104235>.

References

- iPropertyManagement. (2019). *Airbnb statistics (June 2019)*. Available online <https://ipropertymanagement.com/airbnb-statistics/>. (Accessed 10 June 2019).
- Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Editorial: Transformational issues of big data and analytics in networked business. *MIS Quarterly*, 40(4), 807–818.
- Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face images. *Journal of Experimental Psychology: General*, 142(4), 1323–1334.
- Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly*, 26(3), 243–268.
- Baron, R. S., Kerr, N. L., & Miller, N. (1992). *Group process, group decision, group action*. Pacific Grove, CA: Brooks/Cole.
- Bearden, J. N., Murphy, R. O., & Rapoport, A. (2008). Decision biases in revenue management: Some behavioral evidence. *Manufacturing & Service Operations Management*, 10(4), 625–636.
- Bendoly, E. (2013). Real-time feedback and booking behavior in the hospitality industry: Moderating the balance between imperfect judgment and imperfect prescription. *Journal of Operations Management*, 31(1–2), 62–71.
- Benítez-Aurioles, B. (2018). Why are flexible booking policies priced negatively? *Tourism Management*, 67, 312–325.
- Bente, G., Baptist, O., & Leuschner, H. (2012). To buy or not to buy: Influence of seller photos and reputation on buyer trust and purchase behavior. *International Journal of Human-Computer Studies*, 70(1), 1–13.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.
- Bolton, G., Greiner, B., & Ockenfels, A. (2013). Engineering trust: Reciprocity in the production of reputation information. *Management Science*, 59(2), 265–285.
- Bridges, J., & Vázquez, C. (2018). If nearly all Airbnb reviews are positive, does that make them meaningless? *Current Issues in Tourism*, 21(18), 2057–2075.
- Brown, J., & Morgan, J. (2006). Reputation in online auctions: The market for trust. *California Management Review*, 49(1), 61–81.
- Carare, O. (2012). The impact of bestseller rank on demand: Evidence from the app market. *International Economic Review*, 53(3), 717–742.
- Chang, L. J., Doll, B. B., van't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology*, 61(2), 87–105.
- Chattopadhyay, M., & Mitra, S. K. (2019). Do Airbnb host listing attributes influence room pricing homogeneously? *International Journal of Hospitality Management*, 81, 54–64.
- Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10), 1407–1424.
- Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, 19(9), 1508–1519.
- Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, 55, 62–73.
- Feeney, M. (2015). *Is ridesharing safe? Cato institute, policy analysis No. 767*. January 27, 2015. Available at: <https://object.cato.org/sites/cato.org/files/pubs/pdf/pa767.pdf>. (Accessed 10 May 2019).
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144.

- Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., & Goodwill, A. (2018). Pricing in the sharing economy: A hedonic pricing model applied to Airbnb listings. *Journal of Travel & Tourism Marketing*, 35(1), 46–56.
- Gladwell, M. (2000). *The tipping point: How little things can make a big difference*. New York: Little Brown.
- Glick, P., DeMorest, J. A., & Hotze, C. A. (1988). Keeping your distance: Group membership, personal space, and requests for small favors. *Journal of Applied Social Psychology*, 18(4), 315–330. Pt. 2.
- Gray, J. A. (1972). The psychophysiological basis of introversion extraversion: A modification of Eysenck's theory. In V. D. Nebylitsyn, & J. A. Gray (Eds.), *The biological bases of individual behaviour* (pp. 182–205). New York: Academic Press.
- Gregg, D. G., & Walczak, S. (2010). The relationship between website quality, trust and price premiums at online auctions. *Electronic Commerce Research*, 10(1), 1–25.
- Guttentag, D. (2015). Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*, 18(12), 1192–1217.
- Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, 78(5), 837–852.
- Houser, D., & Wooders, J. (2006). Reputation in auctions: Theory, and evidence from eBay. *Journal of Economics and Management Strategy*, 15(2), 353–369.
- Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2007). *Labeled faces in the wild: A database for studying face recognition in unconstrained environments*. Amherst: University of Massachusetts. Technical Report 07-49. Amherst: University of Massachusetts.
- Hu, M., Milner, J., & Wu, J. (2015). Liking and following and the newsvendor: Operations and marketing policies under social influence. *Management Science*, 62(3), 867–879.
- ter Huurne, M., Ronteltap, A., Corten, R., & Buskens, V. (2017). Antecedents of trust in the sharing economy: A systematic review. *Journal of Consumer Behaviour*, 16(6), 485–498.
- Jaeger, B., Slegers, W. W., Evans, A. M., Stel, M., & van Beest, I. (2019). The effects of facial attractiveness and trustworthiness in online peer-to-peer markets. *Journal of Economic Psychology*, 75, 102125.
- Joseph, W. B. (1982). The credibility of physically attractive communicators: A review. *Journal of Advertising*, 11(3), 15–24.
- Juniper Research. (2017). *Sharing economy: Opportunities, impacts and disruptors 2017–2022*. Basingstoke: Juniper Research.
- Kocabiyyikoglu, A., Gogus, C. I., & Gonul, M. S. (2015). Revenue management vs. newsvendor decisions: Does behavioral response mirror normative equivalence? *Production and Operations Management*, 24(5), 750–761.
- Kremer, M., Mantin, B., & Ovchinnikov, A. (2017). Dynamic pricing in the presence of myopic and strategic consumers: Theory and experiment. *Production and Operations Management*, 26(1), 116–133.
- Li, T., Liu, X., Pan, J., & Zhou, G. (2017). The interactive effect of facial appearance and behavior statement on trust belief and trust behavior. *Personality and Individual Differences*, 117, 60–65.
- Luca, M. (2017). Designing online marketplaces: Trust and reputation mechanisms. *Innovation Policy and the Economy*, 17(1), 77–93.
- Maeng, A., & Tanner, R. J. (2013). Construing in a crowd: The effects of social crowding on mental construal. *Journal of Experimental Social Psychology*, 49, 1084–1088.
- Masi, I., Wu, Y., Hassner, T., & Natarajan, P. (2018). Deep face recognition: A survey. In *Proceedings of the Conference on graphics, Patterns and images (SIBGRAPI)*, Parana, Brazil, October.
- Nosko, C., & Tadelis, S. (2015). *The limits of reputation in platform markets: An empirical analysis and field experiment*. National Bureau of Economic Research, Working Paper No. 20830, January. Available at: <https://www.nber.org/papers/w20830.pdf>.
- Park, E. S., & Hinsz, V. B. (2006). Strength and safety in numbers: A theoretical perspective on group influences on approach and avoidance motivation. *Motivation and Emotion*, 30(2), 135–142.
- Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In X. Xie, M. W. Jones, & G. K. L. Tam (Eds.), *Proceedings of the British machine vision conference (BMVC)* (pp. 41.1–41.12). Swansea: BMVA Press, 7–10 September.
- Pavlou, P. A., Liang, H., & Xue, Y. (2007). Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective. *MIS Quarterly*, 31(1), 105–136.
- Ranchordás, S. (2015). Does sharing mean caring: Regulating innovation in the sharing economy. *Minnesota Journal of Law, Science and Technology*, 16(1), 413–475.
- Resnick, P., Zeckhauser, R., Swanson, J., & Lockwood, K. (2006). The value of reputation on eBay: A controlled experiment. *Experimental Economics*, 9(2), 79–101.
- Scheidlinger, S. (1952). *Psychoanalysis and group behavior: A study of Freudian group psychology*. New York: W. W. Norton & Co.
- Stewart, L. H., Ajina, S., Getov, S., Bahrami, B., Todorov, A., & Rees, G. (2012). Unconscious evaluation of faces on social dimensions. *Journal of Experimental Psychology: General*, 141(4), 715.
- Talluri, K. T., & Van Ryzin, G. J. (2006). *The theory and practice of revenue management*. Springer Science & Business Media.
- Teubner, T., Adam, M. T., Camacho, S., & Hassanein, K. (2014). Understanding resource sharing in C2C platforms: The role of picture humanization. In *Proceedings of the 25th Australasian conference on information systems*. December 8–10, Auckland, New Zealand.
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308(5728), 1623–1626.
- Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, 27(6), 813–833.
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, 12(12), 455–460.
- Van't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, 108(3), 796–803.
- Viola, P., & Jones, M. J. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition* (Vol. 1, pp. 511–518).
- Wang, D., & Nicolau, J. L. (2017). Price determinants of sharing economy-based accommodation rental: A study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management*, 62, 120–131.
- Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, 76, 820–838.
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17(7), 592–598.
- Wolf, L., Hassner, T., & Maoz, I. (2011). Face recognition in unconstrained videos with matched background similarity. In *Proceedings of the IEEE computer society conference on computer vision and pattern recognition (CVPR)*, June. Colorado springs, CO. IEEE.
- Yu, M., Saleem, M., & Gonzalez, C. (2014). Developing trust: First impressions and experience. *Journal of Economic Psychology*, 43, 16–29.
- Zekanovic-Korona, L., & Grzunov, J. (2014, May). Evaluation of shared digital economy adoption: Case of Airbnb. In *Proceedings of the 37th international convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 1574–1579). Opatija, Croatia: IEEE. Available online <https://ieeexplore.ieee.org/document/6859816>.
- Zervas, G., Proserpio, D., & Byers, J. (2015). *A first look at online reputation on Airbnb, where every stay is above average* (pp. 1–22). SSRN. Available at <https://ssrn.com/abstract=2554500>.
- Zhang, L., Yan, Q., & Zhang, L. (2018). A computational framework for understanding antecedents of guests' perceived trust towards hosts on Airbnb. *Decision Support Systems*, 115, 105–116.



Stuart J. Barnes is professor of Marketing at King's College London. He holds a first class honours degree in Economics and Geography from University College London and a PhD from Manchester Business School. His research interests include tourism marketing, data analytics, digital innovation, consumer psychology and behaviour, technology acceptance and use, and website evaluation metrics. He has published five books (one a bestseller for Butterworth-Heinemann) and around 200 articles, including those in tourism research, marketing, information systems, and other areas of management.



Samuel Nathan Kirshner is a senior lecturer at the University of New South Wales Business School, where he researches behavioural decision making in supply chain management and technology management problems, and interdisciplinary projects, including analysing technology's ability to enable empathy and ethical decision making. His research has been published in *Manufacturing & Service Operations Management*, *Production and Operations Management*, and the *European Journal of Operational Research*. Sam holds MSc and PhD degrees in Management Science from Queen's University, Canada. He is the co-author of *Business Analytics: A Management Approach*.