International Journal of Information Management 56 (2021) 102265

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Understanding the overvaluation of facial trustworthiness in Airbnb host images

Stuart J. Barnes

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

ARTICLE INFO

Keywords:
Host
Face
Perceived trustworthiness
Dangerous decisions theory
Deep learning

ABSTRACT

Renting a property via a peer-to-peer platform involves a variety of risks. Humans inherently, subconsciously use facial cues as important shortcuts in making assessments about other persons. On property sharing platforms, such as Airbnb, facial cues can be used in a similar fashion alongside reputational information. According to Dangerous Decisions Theory (DDT), intuitive evaluations of trustworthiness based on faces can bias subsequent assessment of an individual, requiring further information sources to make a more balanced assessment. In this study we apply DDT to demonstrate that evaluations based on perceived facial trustworthiness are overvalued; when combined with reputational measures, such as 'super host' status, such assessments are diminished. The study is based on deep learning to classify host faces for a large data set of online accommodation (n = 78,386). The research demonstrates that facial trust cues in online platforms should be treated with caution and must be combined with more objective measures of reputation in order to reduce the effects of overvaluation. The paper concludes with implications for practice and future research.

1. Introduction

Online marketplaces rely on the trust in sellers by buyers to become successful. Transactions in sharing economy platforms contain various risks for buyers, including fulfilment of paid obligations, privacy of data, and accuracy of product or service descriptions (Pavlou et al., 2007). Moreover, sharing economy platforms such as Uber and Airbnb pose further risks for buyers and sellers; for buyers, there are possible safety risks in unfamiliar environments, while sellers may be at risk of property damage. Thus, incidents of assault and robbery have been noted on Uber (Feeney, 2015), while on Airbnb incidence of theft and property damage have been recorded (ter Huurne et al., 2017t), along with breaches of privacy and security, including livestreaming of guests (Zhang et al., 2018). Trust issues are compounded in situations where accommodation is shared with hosts and uncertainties regarding liabilities and legal protection if any injury or damage occurs (Ranchordás, 2015).

To ameliorate risk for buyers, online platforms create reputation systems that enhance transparency and integrate information on participants' historical performance and other elements to help build trust (Luca, 2017). Adding a host photo to a profile can enhance trustworthiness (Bente et al., 2012; Guttentag, 2015), reduce perceived risk and contribute to overall reputation profile, influencing whether a guest chooses to stay at a property (Luca, 2017). Humans make split-second decisions regarding attractiveness and trustworthiness of faces (Willis & Todorov, 2006). However, studies from researchers examining the psychology of decision-making in risky situations, in particular that of judges and jurors in court, have found that such instantaneous evaluations of facial trustworthiness tend to bias decision-making and overall perceptions of credibility (Porter & ten Brinke, 2009). In such situations, facial assessments are overvalued compared to other information (evidence) that is weighed against it. The theory has been empirically tested and supported in legal research (Baker et al., 2016, Porter et al., 2010). We contend that the underlying theory explaining this mechanism, dangerous decisions theory (Porter & ten Brinke, 2009), has value in other areas of human decision-making research and behaviour, such as that of the sharing economy. Thus, through the lens of dangerous decisions theory, we ask the research question: Are perceptions of facial trustworthiness overvalued in the evaluation of hosts on Airbnb?

The overvaluation of host images has implications for both hosts and platform providers. It implies that facial images of a host are the initial, instant source of trustworthiness cues when a potential guest views a listing. As such, they are extremely important as an initial reference point that will influence subsequent information processing about the host. Banerjee and Chua (2020) found that facial cues (happy faces in their study) provided a strong influence on the positive responses of tour guide customers. Similarly, we expect trustworthy faces to influence

E-mail address: stuart.barnes@kcl.ac.uk.

https://doi.org/10.1016/j.ijinfomgt.2020.102265
Received 29 July 2020; Received in revised form 26 October 2020; Accepted 26 October 2020
Available online 5 November 2020

positive evaluations of hosts on Airbnb. However, an overvaluation of host images implies that without other effective trust signals about the host in a profile, the valuation of the host image will play a disproportionate role in overall assessment. This suggests that not only is a trustworthy facial image required as an immediate 'hook', but an effective portfolio of trust cues about a host in the profile is essential. We would expect the overvaluation effect to be ubiquitous in situations where a face is used as a host image.

This research provides three key contributions. First, the study provides an original application of dangerous decisions theory to explain behaviour in information management (and indeed business) research. Second, the research offers a contribution to methodology by advocating the use of deep learning in the development of new image constructs that may be integrated into traditional structural models for analysis. Third, the research findings provide unique evidence that host images are overvalued in customer decision-making on travel accommodation web sites. From a practical perspective, hosts must expend effort not just in building profiles in reputations systems but in conveying the best possible image in additional content that may be processing through heuristics.

The structure of the paper is as follows. In the next section the underlying theory and research model is explained, including hypotheses. In section three, the methods for data collection, processing and analysis are described. The penultimate section presents the results of statistically testing the research model. Finally, in the last section, the findings are discussed and implications for future research and practice are explicated.

2. Theory and hypotheses

This section examines the theory and research model used in the study. The next section examines the theoretical underpinnings of the investigation; this is then followed by the research model and justification of hypotheses.

2.1. Dangerous decisions theory

Our paper focuses on a relatively recent theory developed in legal research. Dangerous decisions theory (DDT) focuses on the potential bias of a decision-maker (a judge in the original research) in overvaluing certain heuristic cues, while discounting other information in making an evaluation (Porter & ten Brinke, 2009). In particular, it focuses on the heuristic appraisal of the face of a subject (a witness or defendant in the original theory) using 'instinct' to provide 'fast and frugal' information regarding credibility.

Fig. 1 outlines the theoretical model for dangerous decisions theory. Here, an initial, automatic judgement of trustworthiness is made rapidly upon seeing an individual's face. Willis and Todorov (2006) found such decisions are most effective after only 100 milliseconds. The initial impression of the subject has an enduring subconscious influence on the manner in which additional evidence regarding the individual is assimilated by the decision-maker. Porter and ten Brinke (2009) refer to this as "'tunnel vision' assimilation of potentially ambiguous or contradictory evidence" concerning the subject. If the subject is initially judged as having low trustworthiness or high trustworthiness, then this assessment 'sticks' and the decision-maker (judge) will seek evidence supporting this preliminary heuristic assessment. Evidence weighed in support of the initial assessment will tend to be overvalued, while contradictory or ambiguous evidence will tend to be undervalued by the decision-maker. Motivation also plays a role in assessment of the subject. High motivation can increase the level of bias in making decisions regarding credibility, particularly for complex tasks (Porter et al., 2007; 2009).

Porter and ten Brinke (2009) point to evidence that "baby-faced individuals receive more lenient judicial outcomes than mature-faced individuals." Among African Americans, those with more Afrocentric features are found to receive harsher sentences (Blair et al., 2004). Further, more attractive defendants are less likely to be found guilty, to receive lower sentences, and be considered less dangerous than unattractive defendants, particularly since perceived attractiveness is positively related to perceived trustworthiness (Bull & Rumsey, 1988; Downs & Lyons, 1991). Wilson and Rule (2015) found that facial trustworthiness predicted extreme sentencing (the death sentence) using a large sample in Florida.

Although DDT was originally designed in the domain of legal research and decision-making by judges and jurors, we contend that the processes explained by the theory are general psychological processes that can potentially be extended to other domains of human behaviour that involve judgement regarding an individual. In this vein, we adopt DDT as an original theory to provide insight into consumer behaviour in online accommodation websites. In our study, the decision-makers are consumers on Airbnb using host images and profile information to make a decision regarding the evaluation of a host (inferred from review score data). The time aspect of the model is consistent with the application of DDT, which posits an initial assessment of facial trustworthiness (such as a judge seeing a defendant), a protracted period of assessment of other information, which then leads to later over-evaluation of facial trustworthiness as related to an overall evaluation of a subject. Although the context of this study is different, we believe that the psychological mechanism of DDT is generalizable between contexts; research has shown that people make split-second assessments of the trustworthiness of faces, regardless of time and context (Willis & Todorov, 2006).

As mentioned in the introduction, we recognise that there are other aspects that will be examined by guests before booking homes. However, we do not aim to capture all elements of the website and these are outside the scope of our assessment of the host via DDT. Indeed, research has shown that trust inferred from profiles (Zhang et al., 2020) and facial trust on Airbnb is positively associated with booking intention (Broeder & Remers, 2018). Such a decision is complex (e.g. see Cao et al., 2020) and involves numerous risks (e.g. ter Huurne et al., 2017t; Zhang et al., 2018), and therefore likely to mirror the same psychological processes.

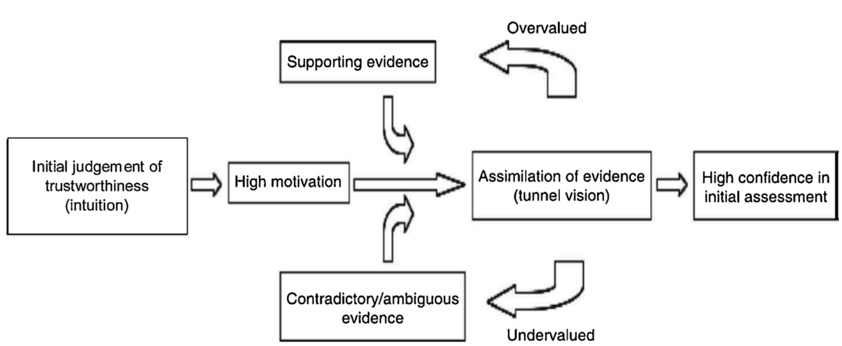

Fig. 1. Dangerous decisions theory (after Porter & ten Brinke, 2009).

2.2. Research model and hypotheses

In this study we apply DDT (Porter & ten Brinke, 2009) to analyse consumer behaviour regarding of the evaluation of Airbnb hosts. The study uses assessment of perceived characteristics of faces, attractiveness and trust, measured using artificial intelligence, summative information regarding the evaluation of a host, overall review rating, and an additional characteristic of the host, super host status. In essence, the research model implies that perceived attractiveness influences perceived trustworthiness, which in turn leads to an inflation in overall rating. The overvaluation of perceived trustworthiness from host faces is tested via the moderating effect of super host status, which we expect to have a negative effect. We also capture the direct effect of super host status on overall rating. Let us briefly examine each of the hypotheses in turn.

Human faces are often used by individuals to make inferences regarding personality (Hassin & Trope, 2000). Such inferences occur within split seconds. Willis and Todorov (2006) found that assessment of attractiveness, trustworthiness, and other characteristics took around 100 milliseconds (for trust this was later found to be only 33 milliseconds; Todorov et al., 2009). Willis and Todorov (2006) further argue that trustworthiness is an automatic processing activity responsible for detecting dangerous stimuli related to human survival, a suggestion supported by Engell et al. (2007).

There is a considerable body of empirical evidence to suggest that perceived attractiveness is positively related to perceived trustworthiness. Research in psychology has found strong support for a close relationship between perceived facial attractiveness and perceived personal trustworthiness (e.g. Gutiérrez-García et al., 2019; Oosterhof & Todorov, 2008; Xu et al., 2012). This has been further supported by evidence from neurophysiological research (e.g. Bzdok et al., 2011; Mende-Siedlecki et al., 2013). The research has been found to be robust across cultural contexts. For example, Xu et al. (2012) conducted a study using Chinese and Caucasian young adults and found that both judged trustworthiness in a similar way. Further they find a significant relationship between attractiveness and trustworthiness. Thus, in this research we posit that:

H1. Perceived attractiveness has a significant positive relationship with perceived trustworthiness.

Our study is based on dangerous decisions theory, which finds a significant relationship between perceived facial trustworthiness and credibility (Porter & ten Brinke, 2009). Yang et al. (2018) operationalise credibility using overall review scores on Airbnb. Previous research has found a significant relationship between perceived trustworthiness and credibility (McGinnies & Ward, 1980). Moreover, the effect of perceived trustworthiness can be more powerful than other drivers of credibility. Wang and Scheinbaum (2018) find that perceived trustworthiness was universally the most important driver of brand credibility. Mao et al. (2020) find that trust inferred from host text profiles is positively associated with booking intention on Airbnb. Ert et al. (2016) assert that facial trust of hosts in Airbnb increases reputation and purchase intention in Airbnb. Numerous studies have found a positive relationship between trustworthiness and price (Ert et al., 2016; Jaeger et al., 2018) and between price and review scores on Airbnb (Chen & Xie, 2017; Gibbs et al., 2018; Teubner et al., 2017; Zhang et al., 2017). Thus, we further hypothesize that:

H2. Perceived trustworthiness is significantly positively related to a host's evaluation rating.

Dangerous decisions theory posits that initial valuations of facial trust are overvalued when weighed against other information sources (Porter & ten Brinke, 2009). Thus, we would expect other information sources to have a negative moderating effect on the relationship between perceived trustworthiness and overall evaluation rating. "Super host" status is an Airbnb badge for the most active and well-reviewed hosts. Biswas et al. (2020) find that super host status moderates the impact of other host attributes on the number of host reviews on Airbnb, which are variously positive or negative according to the type of attribute. We also expect super host status to have a moderating effect. We therefore hypothesize that:

H3. Super host status significantly negatively moderates the relationship between perceived trustworthiness and overall rating.

Research into Airbnb has found the super host status can increase demand for a property (Xie & Mao, 2017). Furthermore, super host status is positively associated with the number of reviews and review score (Liang et al., 2017). Our research follows previous research and posits that:

H4. Super host status has a significant positive relationship with overall rating.

We also add the volume of reviews (average reviews per month) as a control variable in our study. Review volume has been used as a proxy for bookings in recent research (Biswas et al., 2020). The research model and hypotheses are summarised in Fig. 2.

3. Methodology

In this section we summarise the data set, classification using deep learning, and data analysis, including model testing and validation.

3.1. Data set

This study used the public domain dataset provided by insideairbnb.com. We focused on data for 10 popular US cities[1] in the period from 3rd to 19th October 2018. The data used in the study included the overall evaluation rating (out of 100 %), super host status (0 = no, 1 = yes) and host images from 155,008 listings downloaded via MATLAB. The Viola-Jones algorithm with Adaboost (Viola and Jones, 2001) was used to remove any non-faces from the data. This algorithm methodically examines an image with for the face and sub-features and rejects it if facial features are not found. Rejected images included, for example, animals, buildings, or obscured faces. Detected faces were cropped within a bounding box for computational efficiency. Manual screening (by one researcher over two working days) was also performed to detect obvious false positives or negatives; we examined images retained and rejected to ensure that valid faces were not rejected, and non-faces or non-valid faces (e.g. obvious celebrities) were not retained. After processing a total of 96,554 host faces remained.

3.2. Face classification using deep learning

Convolutional neural networks (CNNs) are typically considered the most accurate methods for face recognition (Masi et al., 2018). In this study we use, VGG-Face, developed by the University of Oxford's Visual Geometry Group for face classification. VGG-Face has superior accuracy to other face CNNs, approximately 97.3 % in tests on the YouTube Faces Dataset (Parkhi et al., 2015). Since the images in Airbnb are slightly different to those VGG-Face was trained on (Huang et al., 2007), we fine-tuned the VGG-Face CNN in this study.

We used the MIT Faces dataset (Bainbridge et al., 2013) for retraining the CNN to classify trustworthy/untrustworthy and attractive/unattractive faces in two separate networks. The data set has a diverse set of faces of individuals (n = 2222) by ethnicity, age and gender, considered largely representative of adult US citizens. All faces have ratings for attractiveness (1 = "unattractive" to 5 = "attractive") and trustworthiness (1 = "not at all" to 9 = "extremely"). For retraining

[1] Boston, Chicago, Denver, Los Angeles, New Orleans, New York City, San Diego, San Francisco, Seattle and Washington DC.

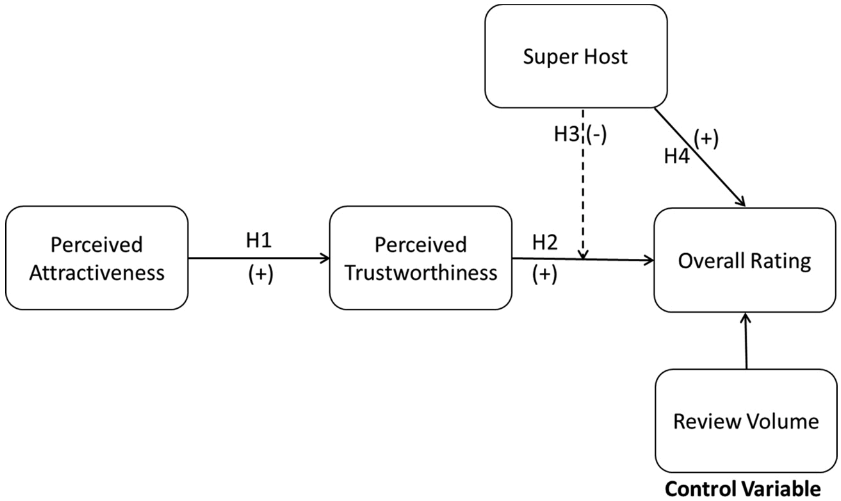

Fig. 2. Research model.

our binary face classifier, we excluded ambiguous faces and focused on those at each end of rating scales. No faces were rated 8 or 9 for trustworthiness and therefore we used faces rated 7–9 on the scale for 'trustworthy' and those rates 1–3 for 'untrustworthy' to retrain the CNN. Similarly, few faces were rated 1 or 5 for attractiveness, and therefore faces rated 4–5 were used for 'attractive' classification and 1–2 for 'unattractive' for retraining the CNN.

The last two layers of the CNN were substituted with those for classifying the new data. We use an increased learning rate factor in the learnable layer to speed up learning in the new learnable layer and froze weights for the first 10 layers in the network (learning rates = 0), to accelerate training and avoid overfitting. Augmentation steps for training images, such as flipping images vertically, randomly translating images by up to 30 pixels, and randomly rescaling images vertically and horizontally by up to 10 %, assisted in avoiding memorization of precise details of training images and overfitting. The number of training epochs (full training cycles on the entire dataset) was set at six, with 78 iterations per epoch. A total of 70 % of the MIT face data was used to train the classifiers; 30 % of the MIT face data was used for validation using the fine-tuned networks. We examined sample image classifications for face validity, and all appeared accurately classified. The networks achieved 99.17 % accuracy for trustworthiness and 91.07 % for attractiveness.

The trained CNNs were applied for the classification of 96,554 host Airbnb faces. The data set was then matched with the numerical data from Airbnb, resulting in 78,386 complete cases (18,168 did not have overall ratings). The descriptive statistics for the variables are shown in Table 1. As we can see, 32.5 % of hosts have the super host status, while the average rating was 94.9 % (s.d. = 7.5 %). Overall, 40.4 % of hosts were classified as attractive, while 79.3 % were classified as trustworthy. The mean number of reviews per month (review volume) was approximately 2 (M = 1.944; S.D. = 2.043).

Table 1
Descriptive statistics for measures.

Measure	Positives	Minimum	Maximum
Perceived Attractiveness	0.404	0	1
Perceived Trustworthiness	0.793	0	1
Super Host	0.325	0	1

Measure	Mean	Median	Min.	Max.	Std. Dev.
Overall Rating	94.921	97	20	100	7.452
Review Volume	1.944	1.2000	0.010	56.050	2.043

3.3. *Data analysis*

We tested the research using partial least squares path modelling (PLSPM) in SmartPLS 3.0 (Ringle et al., 2015). PLSPM is a variance maximization technique for structural equation modelling (SEM) that does not contain distributional assumptions for data samples. Smart-PLS is able to analyse metric variables on different scales in a single model, as is the case in this study (Hair et al., 2014). PLSPM is particularly strong in cases of more complex, predictive models including formative indicators, moderating variables, and single-item measures (Hair et al., 2014). Following the recommendations of Hair et al. (2014; 2019), we choose partial least squares SEM because of the lack of distributional assumptions, the inclusion of single-item formative measurements, and the use of bootstrap moderation analysis. The analysis applied bias corrected, complete bootstrapping via 5000 subsamples.

3.3.1. *Measure validity*

Although single-item constructs have some limitations, they offer practical summative measurement, particularly for unambiguous constructs (Wanous et al., 1997; Bergkvist and Rossiter, 2007). The single-item constructs in thus study are based on explicit measures and therefore appear suitable. However, the single-item constructs used are also formative, so it is not possible to implement standard discriminant validity testing procedures (e.g. cross-loading (Chin, 1998) or via Fornell and Larcker's (1981) method), or to evaluate internal consistency (e.g. Dillon-Goldstein's Rho or Cronbach's Alpha). Notwithstanding, the condition index (Chin, 1998; Duarte and Raposo, 2010) was calculated and found to be below the recommended threshold of 30 for each of our variables, the highest value being 5.166 (see Table 2). This confirmed that multicollinearity was not present in our data set.

To further test for discriminant validity, we used the heterotrait-monotrait test (HTMT; Henseler et al., 2015). The heterotrait-monotrait ratio of correlations was below 0.9 for all variables, the highest being 0.263 (see Table 3). This suggests strong discriminant validity in our data.

Table 2
Condition index analysis in SPSS.

Independents/Dependent	PA	RA	SH	PT	RV
Perceived Attractiveness (PA)	–	5.160	2.402	1.492	1.632
Overall Rating (RA)*	2.365	–	4.764	1.954	2.141
Super Host (SH)	1.416	1.582	–	2.762	2.805
Perceived Trustworthiness (PT)	1.598	2.256	1.511	–	5.166
Review Volume (RV)*	4.714	2.844	1.634	1.311	–

Note: *standardised.

Table 3
Discriminant validity analysis using heterotrait-monotrait ratio (HTMT).

Variable	PA	RA	SH	PT
Perceived Attractiveness (PA)				
Overall Rating (RA)	0.009			
Super Host (SH)	0.055	0.234		
Perceived Trustworthiness (PT)	0.263	0.045	0.048	
Review Volume (RV)	0.039	0.079	0.234	0.007

3.3.2. Model validity

To assess the validity of the structural model, R^2 and Q^2 were examined, since traditional fit measures are considered inappropriate (Hair et al., 2019). SmartPLS was used to analyse model quality using R^2 and Stone-Geisser's Q^2 (Ringle et al., 2015). The values of Q^2 were calculated via blindfolding with an omission distance of 8 and the redundancy technique. We also examine the significance of f^2.

The results are summarised in Table 4. As we can see, the levels of R^2 and Q^2 are 0.1 for perceived trust and 0.1 for overall rating. These levels are considered acceptable and have moderate to small predictive validity (Falk and Miller, 1992). All f^2 values were found to be significant at $p < .001$.

4. Results

In this section, we provide the results of statistically testing our research model and hypotheses. Table 5 provides the results of testing the research model using complete bootstrapping in SmartPLS. Table 5 provides the path coefficients and significance for the four the hypotheses tested in our research model. The results demonstrate a strongly significant positive relationship between the perceived attractiveness and perceived trust ($\beta = 0.263$, $t = 90.751$, $p < .001$), supporting H1 – *perceived attractiveness has a significant positive relationship with perceived trustworthiness*. Similarly, the sample demonstrates a very strong positive relationship between perceived trustworthiness and host evaluating rating ($\beta = 0.033$, $t = 8.966$, $p < .001$), demonstrating empirical support for H2, whereby *perceived trustworthiness is significantly positively related to a host's evaluation rating.*

The negative moderating effect of super host status on the relationship between perceived trustworthiness and overall rating was significant (β=-0.011, $t = 4.068$, $p < .001$), providing evidence in support of H3, that *super host status significantly negatively moderates the relationship between perceived trustworthiness and overall rating.* More details on the moderation bootstrap analysis are provided in Table 6. As we can see, the bias-corrected 95 % confidence interval does not contain zero, confirming the strength of the moderation effect. The moderating effect of super host status is further examined in Fig. 3. As we can see, super host status results in higher mean overall ratings for hosts with both high (97.5 %) and low trustworthiness (97.2 %). However, the improvement in mean overall rating is higher for hosts with low perceived trustworthiness (4.1 %) compared to hosts with a high perceived trustworthiness (3.6 %), demonstrating the negative moderating effect.

Finally, we find that super host status had a highly significant direct positive relationship with overall rating ($\beta = 0.233$, $t = 94.026$, $p < .001$), supporting H4 – *super host status has a significant positive relationship with overall rating.* The control variable, review volume, did not have a significant relationship with overall rating ($p = .988$). Fig. 4 summarises the results of testing our research model.

Table 4
Predictive validity analysis.

Variable	SSO	SSE	Q^2	R^2	f^2 sig.
Overall Rating	78386.0	74027.5	0.1	0.1	$p < .001$
Perceived Trust	78386.0	72997.1	0.1	0.1	$p < .001$

Table 5
Results of Bootstrapping Analysis.

Relationship and Hypothesis Tested	Beta	t-value	p-value
H1. Perceived Attractiveness -> Perceived Trustworthiness	0.263	90.751	<.001
H2. Perceived Trustworthiness -> Overall Rating	0.033	8.966	<.001
H3. Super Host Moderator	−0.011	4.068	<.001
H4. Super Host -> Overall Rating	0.233	94.026	<.001
Control Variable. Review Volume -> Overall Rating	0.000	0.015	0.988

Table 6
Bootstrap Moderation Analysis: Super Host Status.

Original Sample	Sample Mean	Standard Deviation	t-value	2.50 %	97.50 %	Bias
−0.011	−0.011	0.003	4.068	−0.016	−0.006	0.000

5. Discussion and conclusions

5.1. Discussion

Deciding on accommodation through sharing economy platforms can be both risky and complex (Cao et al., 2020; ter Huurne et al., 2017t; Zhang et al., 2018). Through the original lens of dangerous decisions theory, this research confirms the over-importance of host profile images in conveying perceived trust and building reputation (through review scores) on sharing economy platforms. The study supports the assertion that perceived attractiveness is related to perceived trustworthiness (H1), a finding supported in previous psychological, neurophysiological and legal research (Bull & Rumsey, 1988; Downs & Lyons, 1991; Oosterhof & Todorov, 2008; Xu et al., 2012; Mende-Siedlecki et al., 2013). Dangerous decisions theory posits that value perceptions from host images will be overvalued and have a significant influence on host assessments from review ratings (Porter & ten Brinke, 2009), a finding supported by the study (H2). This confirms previous research asserting a significant relationship between perceived trustworthiness and credibility (McGinnies & Ward, 1980; Wang & Scheinbaum, 2018), or more specifically between perceived trustworthiness and review rating, often via price (Chen & Xie, 2017; Jaeger et al., 2018; Zhang et al., 2017).

The overvaluation effect from DDT is demonstrated by the negative moderating effect of other reputational information, in this case super host status (support of H3). This provides a clear answer to our research question and an original finding in information management (and business) research. Finally, our research confirms the positive relationship between super host status and review score (H4), supporting the findings of Liang et al. (2017). Originally developed in legal research, this paper demonstrates that DDT has value for explaining behaviour in other areas, such as online consumer behaviour in the sharing economy.

One interesting observation from the research is that the improvement in mean host evaluation via super host status is greater for hosts with low perceived trustworthiness as compared to hosts with a high perceived trustworthiness. This suggests that although perceived trustworthiness is an important first trust signal for a consumer, super host status is more powerful overall and provides a means of levelling the playing field. Thus, even if facial images display poor trustworthiness, hosts can compensate by working hard to obtain super host status to build a stronger overall evaluation.

5.2. Implications for theory and research

The research confirms the importance of trustworthy host images in driving positive perceptions of accommodation services delivered to customers on Airbnb, measured through review scores. Review scores

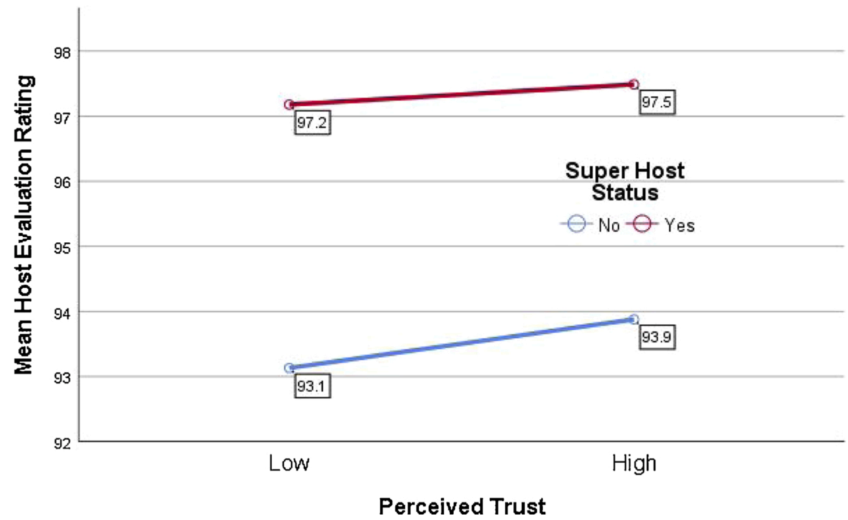

Fig. 3. The moderating effect of super host status.

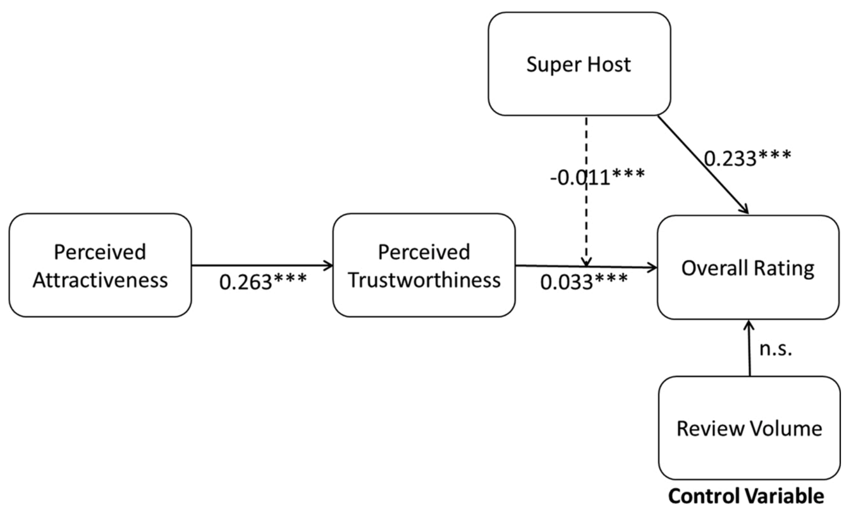

Fig. 4. Results of model testing.

and other reputational information, such as super host status, are key information sources for consumers deciding to book accommodation on sharing economy platforms. Facial images are quickly processed by consumers and provide instantaneous information on perceived trustworthiness (and perceived attractiveness, one of its determinants). Such facial cues are biased, and these first impressions have a disproportionate affect; perceptions of facial trust 'stick' and are overvalued by consumers. Thus, high quality reputational information from other sources is needed to temper these initial impressions (for example super host status).

This study has only scratched the surface of the application of DDT in the sharing economy. In the context of travel and tourism websites, it would be desirable to extend the research model to understand its applicability with respect to other relevant signals from the characteristics of the host or features from the Airbnb profile of the host. For example, it would be possible to examine how trust perceptions are affected by customer characteristics such as race, country of origin, cultural values and ethnicity (Gupta et al., 2019; Kakar et al., 2018) and gender (Su & Mattila, 2019). It would also be helpful to examine the moderating effect within DDT of other information provided by the host. For example, a host description has been found to be important in performance (Liang et al., 2020). Specific elements in host self-description that are particular important in influencing perceived trust and purchasing include readability, perspective, family-orientation, and openness (Zhang et al., 2020). Other aspects such as host interaction and communication have been found to influence satisfaction (Xu, 2020). Similarly, it may be interesting to disentangle host and property signals, including depth and volume of property descriptions (Liang et al., 2020), access, amenities, external environment, and prosperity of a locality (Wang & Nicolau, 2017; Xu, 2020)

This research demonstrates the value of dangerous decisions theory in understanding aspects of the emerging sharing economy and the usefulness of CNN in providing facial image classifications to develop variables for social science research. DDT provides a new theory in information management research that could be applied in many studies where facial images are involved, such as social media. For example, it could also potentially be applied in understanding persuasive communication in social media, including influencer marketing, viral videos on TikTok, or the effectiveness of fabricated news. It is hoped that this investigation encourages researchers to consider other areas of application of DDT and to integrate facial image analytics using CNN into their research models in the future.

5.3. Implications for management

The overvaluation of host images has considerable implications for both hosts and platform providers. It implies that facial images of a host are the initial, instant source of trustworthiness cues when a potential guest views a listing. As such, they are extremely important as an initial reference point that will influence subsequent information processing about the host. Moreover, attractive facial images will be perceived as more trustworthy. This implies, at a simple level, that curating appropriate host images that are both trustworthy and attractive is important, but the implications go much deeper than this. Time invested in delivering the best impression from a host image may reap dividends in terms of the improvements in review scores obtained, indirectly leading to more customer patronage and revenues. However, as mentioned in Section 5.3, super host status appears more effective in leveraging improvements in overall evaluation for hosts. Thus, even if host images display low perceived trustworthiness, hosts can compensate significantly by striving for super host status. This could potentially be because super host status is seen as a more reliable validation of trustworthiness.

An overvaluation of host images implies that without other effective trust signals about the host in a profile, the valuation of the host image will play a disproportionate role in overall assessment. Not only must online platforms provide appropriate trust mechanisms, they must be populated effectively by hosts, to create a portfolio of trust cues about a host and their property in the profile to support initial perceptions of facial trustworthiness. Such trust cues include clear, open, honest and informative information about the host and the property, and relevant travel information.

5.4. Limitations

This study has several limitations. First, although we have a very large data set, it is limited to 10 US cities, albeit throughout the country. Other US locations may have different host populations and it would therefore be apposite to test the model using a larger data set of a broader number of cities and other geographic locations, including towns and villages. Second, our study is limited to the US context and it would be desirable to test the research model in other country contexts (which would require appropriate training and validation data sources).

Third, there are some potential limitations of the training dataset. Although broadly representative of adult US citizens, including age and ethnicity, the training dataset is limited in both size and granularity, with few ratings on the extreme poles. Thus, we focused on training a dual classification CNN model. A larger amount of more varied training could allow greater sensitivity and improve the accuracy of the classification model.

Finally, we recognise that the focus of our study is limited to perceived facial trustworthiness within the scope of DDT. Other crucial demographic and anthropological factors such as race, country of origin, ethnicity, gender, prosperity/median income of a locality (and whether it matches with that of the prospective guest), can also affect the perceived trustworthiness and contribute to a more comprehensive model (Kakar et al., 2018; Su & Mattila, 2019; Wang & Nicolau, 2017). These provide potential directions for developing the research in the future.

Author statement

This is a single-authored paper. All aspects of the research were completed by the author.

References

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face images. *Journal of Experimental Psychology General, 142*(4), 1323–1334.
Baker, A., Porter, S., ten Brinke, L., & Mundy, C. (2016). Seeing is believing: Observer perceptions of trait trustworthiness predict perceptions of honesty in high-stakes emotional appeals. *Psychology Crime and Law, 22*(9), 817–831.
Banerjee, S., & Chua, A. Y. (2020). How alluring is the online profile of tour guides? *Annals of Tourism Research, 81*, Article 102887. https://doi.org/10.1016/j.annals.2020.102887.
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.
Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research, 44*, 175–184.
Biswas, B., Sengupta, P., & Chatterjee, D. (2020). Examining the determinants of the count of customer reviews in peer-to-peer home-sharing platforms using clustering and count regression techniques. *Decision Support Systems, 135*, Article 113324. https://doi.org/10.1016/j.dss.2020.113324.
Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of afrocentric facial features in criminal sentencing. *Psychological Science, 15*, 674–679.
Broeder, P., & Remers, E. (2018). Eye contact and trust online: The effect of profile pictures on Airbnb booking. In *Proceedings of the 12th IEEE International Conference on Application of Information and Communication Technologies (AICT)* (pp. 1–4).
Bull, R., & Rumsey, N. (1988). *The social psychology of facial appearance*. New York: Springer.
Bzdok, D., Langner, R., Caspers, S., Kurz, F., Habel, U., Zilles, K., … Eickhoff, S. B. (2011). ALE meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain Structure & Function, 215*, 209–223.
Cao, Q., Liu, X., Wang, Z., Zhang, S., & Wu, J. (2020). Recommendation decision-making algorithm for sharing accommodation using probabilistic hesitant fuzzy sets and bipartite network projection. *Complex & Intelligent Systems, 6*, 431–445.
Chen, Y., & Xie, K. (2017). Consumer valuation of Airbnb listings: A hedonic pricing approach. *International Journal of Contemporary Hospitality Management, 29*(9), 2405–2424.
Chin, W. W. (1998). The partial least squares approach for structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 236–295). London: Lawrence Erlbaum.
Downs, A. C., & Lyons, P. M. (1991). Natural observations of the links between attractiveness and initial legal judgments. *Personality & Social Psychology Bulletin, 17*, 541–547.
Duarte, P. A. O., & Raposo, M. L. B. (2010). A PLS model to study brand preference: An application to the mobile phone market. In V. Esposito Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares: Concepts, methods and applications* (pp. 449–485). Berlin: Springer-Verlag.
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.
Falk, R. F., & Miller, N. B. (1992). *A primer for Soft Modeling*. Akron: University of Akron Press.
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 [last accessed 10 May 2019].
Fornell, C., & Larcker, F. D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.
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.
Gupta, M., Esmaeilzadeh, P., Uz, I., & Tennant, V. M. (2019). The effects of national cultural values on individuals' intention to participate in peer-to-peer sharing economy. *Journal of Business Research, 97*, 20–29.
Gutiérrez-García, A., Beltrán, D., & Calvo, M. G. (2019). Facial attractiveness impressions precede trustworthiness inferences: Lower detection thresholds and faster decision latencies. *Cognition & Emotion, 33*(2), 378–385.
Guttentag, D. (2015). Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism, 18*(12), 1192–1217.
Hair, J. F., Hult, G. M. T., Ringle, C. M., & Sarstedt, M. (2014). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Thousand Oaks: Sage.
Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24.
Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology, 78*(5), 837.
Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135.
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.
Jaeger, B., Sleegers, W. W., Evans, A. M., Stel, M., & van Beest, I. (2018). The effects of facial attractiveness and trustworthiness in online peer-to-peer markets. *Journal of Economic Psychology, 75*(Part A), Article 102125.
Kakar, V., Voelz, J., Wu, J., & Franco, J. (2018). The visible host: Does race guide Airbnb rental rates in San Francisco? *Journal of Housing Economics, 40*, 25–40.
Liang, S., Schuckert, M., Law, R., & Chen, C. C. (2017). "Be a 'superhost'": The importance of badge systems for peer-to-peer rental accommodations. *Tourism Management, 60*, 454–465.
Liang, S., Schuckert, M., Law, R., & Chen, C. C. (2020). The importance of marketer-generated content to peer-to-peer property rental platforms: Evidence from Airbnb.

International Journal of Hospitality Management, 84, Article 102329. https://doi.org/10.1016/j.ijhm.2019.102329.

Luca, M. (2017). Designing online marketplaces: Trust and reputation mechanisms. *NBER/Innovation Policy and the Economy, 17*(1), 77–93.

Mao, Z., Jones, M. F., Li, M., Wei, W., & Lyu, J. (2020). Sleeping in a stranger's home: A trust formation model for Airbnb. *Journal of Hospitality and Tourism Management, 42*, 67–76.

Masi, I., Wu, Y., Hassner, T., & Natarajan, P. (2018). Deep face recognition: A survey. *Proceedings of the Conference on Graphics, Patterns and Images (SIBGRAPI)*.

McGinnies, E., & Ward, C. D. (1980). Better liked than right: Trustworthiness and expertise as factors in credibility. *Personality & Social Psychology Bulletin, 6*(3), 467–472.

Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2013). The social evaluation of faces: A meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience, 8*, 285–299.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*, 11087–11092.

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.

Porter, S., & ten Brinke, L. (2009). Dangerous decisions: A theoretical framework for understanding how judges assess credibility in the courtroom. *Legal and Criminological Psychology, 14*, 119–134.

Porter, S., McCabe, S., Woodworth, M., & Peace, K. A. (2007). Genius is 1% inspiration and 99% perspiration? Or is it? an investigation of the effects of motivation and feedback on deception detection. *Legal and Criminological Psychology, 12*, 297–309.

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology Crime and Law, 16*(6), 477–491.

Ranchordás, S. (2015). Does sharing mean caring: Regulating innovation in the sharing economy. *Minnesota Journal of Law, Science & Technology, 16*(1), 413–475.

Ringle, C. M., Wende, S., & Becker, J.-M. (2015). *SmartPLS 3*. Boenningstedt: SmartPLS GmbH. http://www.smartpls.com.

Su, N., & Mattila, A. S. (2019). Does gender bias exist? The impact of gender congruity on consumer's Airbnb booking intention and the mediating role of trust. *International Journal of Hospitality Management*. , Article 102405. https://doi.org/10.1016/j.ijhm.2019.102405.

ter Huurne, M., Ronteltap, A., Corten, R., & Buskens, V. (2017t). Antecedents of trust in the sharing economy: A systematic review. *Journal of Consumer Behaviour, 16*(6), 485–498.

Teubner, T., Hawlitschek, F., & Dann, D. (2017). Price determinants on Airbnb: How reputation pays off in the sharing economy. *Journal of Self-Governance and Management Economics, 5*(4), 53–80.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*(6), 813–833.

Viola, P., & Jones, M. J. (2001). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1*, 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.

Wang, S. W., & Scheinbaum, A. C. (2018). Enhancing brand credibility via celebrity endorsement. *Journal of Advertising Research, 58*(1), 16–32.

Wanous, J. P., Reichers, A., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *The Journal of Applied Psychology, 82*(2), 247–252.

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.

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science, 26*(8), 1325–1331.

Xie, K., & Mao, Z. (2017). The impacts of quality and quantity attributes of Airbnb hosts on listing performance. *International Journal of Contemporary Hospitality Management, 29*(9), 2240–2260.

Xu, X. (2020). How do consumers in the sharing economy value sharing? Evidence from online reviews. *Decision Support Systems, 128*, Article 113162. https://doi.org/10.1016/j.dss.2019.113162.

Xu, F., Wu, D., Toriyama, R., Ma, F., Itakura, S., & Lee, K. (2012). Similarities and differences in Chinese and Causasian adults' use of facial cues for trustworthiness judgments. *PloS One, 7*(4), Article e34859. https://doi.org/10.1371/journal.pone.0034859.

Yang, S.-B., Lee, H., Lee, K., & Koo, C. (2018). The application of Aristotle's rhetorical theory to the sharing economy: An empirical study of Airbnb. *Journal of Travel & Tourism Marketing, 35*(7), 938–957.

Zhang, Z., Chen, R. J., Han, L. D., & Yang, L. (2017). Key factors affecting the price of Airbnb listings: A geographically weighted approach. *Sustainability, 9*, 1–13.

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.

Zhang, L., Yan, Q., & Zhang, L. (2020). A text analytics framework for understanding the relationships among host self-description, trust perception and purchase behavior on Airbnb. *Decision Support Systems*. , Article 113288. https://doi.org/10.1016/j.dss.2020.113288.