First Impression: informative visual cues on digital platform

by

Xiaohang (Flora) Feng

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Tepper School of Business, Carnegie Mellon University

First Impression: informative visual cues on digital platform

Xiaohang (Flora) Feng

Doctor of Philosophy, 2025

Tepper School of Business, Carnegie Mellon University

Abstract

This dissertation shows how first impressions, via faces, image style, and at-a-glance informational signals, shape choice, prices, and market structure on digital platforms. I argue these cues are measurable, causal, and governable. The first study develops an explainable computer-vision method to score celebrity visual potential (CVP) from 11 facial features tied to charisma-related inferences. Trained on a large face corpus and validated in controlled experiments and field data (Instagram, LinkedIn), the model predicts celebrity status with high accuracy, aligns with human judgments, and adds signal beyond attractiveness or typicality. The second study quantifies visual uniqueness in peer-to-peer marketplaces using an unsupervised, psychology-guided representation of 481,747 listing images with multiple human validations (including eye tracking). Demand on Airbnb follows an inverted-U: moderate distinctiveness boosts bookings, while excessive distinctiveness backfires; gains are larger when hosts signal responsiveness and quality. The third study examines Amazon's Climate Pledge Friendly badge. A three-stage price–badging game and large-scale causal evidence show that a unified sustainability signal increases demand and prices and reduces market concentration under realistic consumer mixes. Overall, the dissertation delivers interpretable measures of glance-level

cues, credible evidence of their market impact, and a managerial playbook that treats first-impression design as a tunable policy under clear ethical guardrails.

CONTENTS

CHAPTER I: INTRODUCTION

Consumers make many platform decisions, whom to follow, which listing to click, what product to buy, within a few seconds and often on a small screen. In these glance-level moments, first-impression cues do most of the work. Some cues are visual, such as a face in a profile photo or the distinctiveness of a listing image. Others are informational at a glance, such as a sustainability badge that compresses complex product attributes into a single, credible signal. This dissertation examines how such cues shape instant inferences, steer choices and willingness to pay, and, when repeated at scale, shift reviews, pricing, and market structure. It also studies how platforms and sellers can govern these effects through design and policy.

By "first-impression cues," I mean high-salience signals processed rapidly with limited deliberation, typically at the thumbnail or search-results layer where attention is scarce and time is short. The conceptual backbone is a common chain: cue, inference, choice and price, market dynamics, and governance. A face, a distinctive image, or a green badge triggers immediate beliefs about charisma, quality, or sustainability; these beliefs move clicks and conversions; accumulated over millions of interactions, they alter concentration and competitive balance; and platform and seller decisions, who appears, how images look, which items are badged, and at what coverage, govern the magnitude and distribution of these effects.

The three core studies follow this logic in sequence. Chapter II develops and validates an explainable computer-vision method to score "celebrity visual potential" from facial features that correlate with charisma-related trait inferences. The metric predicts celebrity status, aligns with human judgments, and adds predictive power beyond attractiveness or typicality, showing that

who appears in the first frame can measurably shift persuasion. Chapter III builds an unsupervised, interpretable model of visual uniqueness for peer-to-peer listings and links it to demand, revealing an inverted-U pattern in which moderate distinctiveness helps but excessive distinctiveness hurts, especially conditional on response rate and ratings. Chapter IV turns to an informational cue, Amazon's Climate Pledge Friendly badge, and combines a theoretical pricing–badging game with causal evidence to show that a unified sustainability signal raises demand and prices and reduces market concentration when pro-badge consumers are sufficiently prevalent.

Across chapters, the dissertation asks three questions: how to measure first-impression cues in ways that are accurate, interpretable, and actionable; what their causal impacts are on choice, price, reviews, and concentration; and how platforms and sellers should set coverage, select endorsers, and tune image style to balance revenue, consumer welfare, and fairness. Methodologically, the work combines explainable computer vision to derive interpretable features, large-scale observational causal inference suited to staggered and reversible treatments, and economic modeling to connect micro-responses to platform-level policy.

The contributions are threefold. Conceptually, the dissertation translates thin-slice social judgments and perceptual distinctiveness into operational metrics that decision makers can use. Empirically, it documents that first-impression cues have sizable and systematic effects on demand and price and, at scale, on market structure. Managerially, it proposes a governance playbook that treats endorser selection, visual distinctiveness, and badge policy, especially coverage, as tunable levers subject to monitoring and audit. Given the potential for bias or misinterpretation, the chapters also articulate ethical guardrails, including fairness considerations in facial modeling and transparency around labeling.

The roadmap is straightforward. Chapter II introduces the CVP construct and validation studies. Chapter III develops the visual-uniqueness pipeline and estimates its non-linear demand effects and moderators. Chapter IV analyzes the CPF badge to link first-impression signals to pricing and competition, bridging individual behavior and market structure. A concluding chapter synthesizes these results into practical guidance for platforms and sellers. In sum, the dissertation argues that the first-impression effect is measurable, causal, and governable: small signals at first glance, faces, visual style, and badges, produce large, predictable shifts in how digital markets function.

CHAPTER II: AN AI METHOD TO SCORE CELEBRITY VISUAL POTENTIAL

ABSTRACT

It has long been a mantra of marketing practice that, particularly in low-involvement situations, spokespeople should be physically attractive. This paper suggests there is a higher probability of gaining fame and influence (i.e., celebrity potential) than is captured by attractiveness or typicality. The authors identify 11 facial features that may predict celebrity potential by virtue of their purported relationship with charisma and resulting personality trait inferences. Using machine learning methods and a sample of 22,000 faces, the authors calculate the direction and strength of the correlation of each feature with celebrity potential. The model is 95.92% accurate in predicting whether a given face belongs to a celebrity or noncelebrity, and it allows calculating a *celebrity visual potential* (CVP) metric for any face. Two controlled experiments and two studies using photographs of faces of Instagram and LinkedIn users further validate that the model-generated CVP is consistent with human-rated CVP, showing predictive power above and beyond facial typicality and averageness. This paper challenges prior assumptions about the importance of attractiveness in spokesperson choice, offers a useful additional metric for marketers, and provides novel insights about the relative importance of various inferred personality traits for celebrity potential.

Keywords: Celebrity Visual Potential, Facial Features, Personality Traits, Deep Learning, Explainable Artificial Intelligence (XAI).

INTRODUCTION

For brands, the increasingly visual nature and democratization of influence offer both an opportunity and a challenge. On the one hand, brands have a virtually infinite supply of potential spokespeople, influencers, and bloggers from whom to draw. They may even simply generate virtual spokespeople (Hwang, Liu, and Srinivasan 2021), further increasing their choice set of various spokespeople. On the other hand, there is little beyond the time-honored suggestion that celebrities should be physically attractive or familiar (Kahle and Homer 1985; Faerber et al. 2016) to help marketers predict the likelihood that a given visually depicted influencer will have enough charisma to drive the inference of positive personality traits, and with it, to gain celebrity. Further, we know little about whether a celebrity can be predicted by the same visual traits across social media platforms, or if a corporate "celebrity" (e.g., a C-suite executive) has the potential for celebrity rooted in different features than an influencer on Instagram.

Celebrity itself is a comprehensive quality that can be judged from both (inner) personality traits and (outer) physical appearance. In this paper, we focus on the visual aspect of celebrity (Moraes et al. 2019; Verčič and Verčič 2011), as manifest in facial features, which has been studied widely (Troncoso and Luo 2023; Banker et al. 2023; Chen et al. 2023). Critically, we propose the construct of *celebrity visual potential* (CVP) as a metric that can be developed based on past theories related to charisma, facial features, and inferred personality traits, and through machine learning models and experimental validation. CVP can play a crucial role in different contexts. If a high CVP salesforce accomplishes significantly higher sales, companies can direct training and support resources more effectively to those with lower CVP, creating a more level playing field. Also, physician detailing is an expensive and important marketing activity.

Salesforce with higher CVP, potentially having more persuasion capabilities, may accelerate physicians' education and prescription of such products.

To assess the importance of CVP, we identify 11 facial features that past literature suggests may drive this metric, based on their relationship with inferred charisma, a central characteristic in celebrity. Six of these (i.e., dominance, warmth, competence, trustworthiness, generosity, and aggressiveness) are inferred personality traits[1] that may mediate the connections between facial features and celebrity potential (Gray, Ward, and Norton 2014; Keating 2002, 2011). We then develop a high-performing, generalizable, and scalable model informed by a unique dataset of 6,000 celebrity and 6,000 noncelebrity images. The model's output is the *CVP score* (CVP), or the probability that the input face is a celebrity. Notably, although the input to CVP is facial features, it is not equivalent to physical appearance or attractiveness, because CVP is also a proxy for persuasiveness. By measuring the direction and ranking the strength of the correlation between CVP and each hypothesized facial feature, we identify the stronger and the weaker drivers of CVP. To bolster our causal argument, we then report one controlled experiment and a conceptual replicate that suggest the convergence of model-generated CVP and consumers' predictions of celebrity potential. These studies also help us understand which previously identified inferred personality traits explain the effects of facial characteristics on CVP. Additionally, we show the power of our model across demographic groups and platforms. Across our data, CVP's explanatory power is above and beyond predictions that would be made based on facial attractiveness alone.

This paper contributes large-scale empirical evidence for the relationship between 11 different facial features and CVP. Although various facial attributes and personality traits have

[1] A personality trait is an internal characteristic (Cattell 1950) excluding attractiveness, though it may mediate the relationships between facial features and CVP (Berscheid and Hatfield 1978; Griffin and Langlois 2006).

been studied in marketing and economics, most studies have limited generalizability because they used small samples or human judgment to rate inferred personality traits (Gangestad et al. 2004; Penton-Voak et al. 1999). Further, prior work has usually considered only one or a small group of facial features at a time; this approach makes it difficult to estimate the overall effect of a composite face on visually imputed charisma, a key factor in celebrity (Stirrat and Perrett 2012). In addition, in some cases, facial features have been suggested to be associated with two inferred personality traits that could have countervailing effects on charisma and thus celebrity potential. For example, in one study a feature might be associated with dominance, a positive predictor of charisma (Wong, Ormiston, and Haselhuhn 2011) and in another study, with aggressiveness, a negative predictor of charisma (Carré, McCormick, and Mondloch 2009). Using machine learning to understand these relationships allows us to conduct a larger-scale analysis that is not only generalizable but also offers rich insights into the relationships between these facial features, as well as providing novel insights into the role of various personality inferences in determining CVP. Finally, we empirically demonstrate that attractiveness, while important, does not tell the whole story of CVP. Thus, we extend prior research on people's responses to others' facial characteristics, providing a more complete picture than presented in prior work (Alley and Cunningham 1991; Baker and Gilbert 1977; Griffin and Langlois 2006).

Methodologically, we join other recent studies (e.g., Zhou et al. 2021) that improve the interpretability of machine learning predictions. Deep learning distinguishes itself from parametric models, such as regression, by accommodating nonlinear, complicated variable relationships with high flexibility in an entirely data-driven way, leading to much higher prediction accuracy than regression models (Schulz et al., 2020). Yet, the "black-box" nature of deep learning makes model predictions hard to interpret. However, we employ interpretable

machine learning techniques to construct a metric that allows identifying the facial features that most affect CVP, leading to clear practical and managerial implications.

Our results have significant managerial implications for marketing, media, entertainment, and business. Marketing researchers are interested in exploring key image features related to economic value (Grewal, Gupta, and Hamilton 2021; Li and Xie 2019; Zhang and Luo 2022; Zhang, Mehta, Singh, and Srinivasan 2021; Zhou, Lu, and Ding 2020), among which human face is a popular topic (Hartmann et al. 2021; Hu and Ma 2021; Kachur et al. 2020; Peng et al. 2022; Wang and Kosinski 2018; Zhou et al. 2021). The CVP metric may be fruitfully used in such efforts as a reference in career choice, as a control variable in randomized experiments that aim to improve persuasiveness or celebrity influence, or as a factor to examine whether human resource professionals excessively use facial features in recruiting. Further, our model may help optimize the design or enhancement of AI-generated virtual personas, such as fashion models, digital influencers, video game or movie characters, and figures in advertising campaigns. Additionally, CVP can be used as one of the indexes for screening and recruiting salesforce, a costly activity for companies such as big pharmaceuticals. However, there are also a couple of noteworthy aspects of CVP that merit careful consideration. We close with a discussion of how CVP can be further examined and applied in theoretical, practical, and policy contexts.

THEORETICAL FRAMEWORK

Celebrity and Charisma

What determines celebrity? A substantial literature would argue that when it comes to visual representation, the most important characteristic of a celebrity is their physical attractiveness. It has been shown to facilitate attitude change in communications and interactions (Baker and Gilbert 1977; Chaiken 1979; Bersheid and Walster 1974). The physical attractiveness of

celebrity endorsers often has an obvious impact on product sales (Kahle and Homer 1985), since even a two-second glance at an advertisement is enough to leave a meaningful impression (Kahle, Kulka, and Klingel 1980).

However, we argue that celebrity potential extends far beyond attractiveness. Though many bases of celebrity have been proposed, psychology literature suggests celebrities across domains share the trait of charisma as a central feature (Potts 2009). Charisma is defined as "a leader's moral conviction, need for power, and ability to transfer an idealized vision to followers" (Antonakis, Fenley, and Liechti 2012; Conger and Kanungo 1987; Weber, Henderson, and Parsons 1947). Many scholars argue that charisma plays a key role in the formulation and maintenance of celebrity power, image, and status (Marshall 1997; Rojek 2001; Alexander 2010). Charisma, in turn, has been argued to be driven by a combination of personality traits, including a sense of power or dominance, trustworthiness, competence, aggressiveness, warmth, and generosity, as well as physical attractiveness (House and Howell 1992; House, Spangler, and Woycke 1991; Moraes et al. 2019; Rein et al. 2006), though predictions of attractiveness vary. Specifically, charisma is positively associated with dominance (Keating 2002); warmth, which leads to approach behaviors (Keating 2011); generosity, which is regarded as an important virtue for good leadership and social recognition (Beck 2012; Winterich, Mittal, and Aquino 2013); and competence, which elicits positive emotions (Avolio and Bass 1988; Gray, Ward, and Norton 2014). By contrast, aggressiveness decreases charisma by reducing a person's ability to arouse positive emotions in other people (Buss and Perry 1992; Costa, McCrae, and Dembroski 1989). The existing literature has diverging conclusions on the impact of physical attractiveness on charisma. For example, Riggio (1987) argues that the effect of attractiveness on charisma is

small. Others believe that attractiveness influences first impressions and correlates with positive qualities (Dion, Berscheid, and Walster 1972; Griffin and Langlois 2006).

In summary, charisma is a broad concept and can be explained by combinations of traits such as success, popularity, and attractiveness, but it differs from each individual trait. For instance, success can be measured by income or company hierarchy, leading to individual identity and satisfaction (Goffman 1959; Arthur et al. 2005); however, charisma differs from success in that charisma emphasizes interpersonal influence. Popularity overlaps with celebrity potential in terms of social acceptance and the quality of being well-liked (Eder 1985); however, unlike popularity, charisma includes the ability to elicit trust and respect. Attractiveness involves external, short-term physical charm (Swami and Furnham 2008), while charisma extends to a multifaceted, long-term influence on the perceiver's mind and emotions (Chase 2016).

Facial Features and CVP

However, prior literature fails to determine how people might infer success, popularity, and attractiveness from visual stimuli, or when one inferred trait might override another in generating charisma and thus potential celebrity. Therefore, we draw from the literature in psychology, sociology, economics, and behavioral marketing and identify 11 facial features relevant to inferring these personality traits. We thus determine their relationship with charisma and, by extension, to CVP. Table 1 summarizes the possible theoretical relationships between CVP and the 11 facial features we identified, with the six personality traits as mediators.

Table 1. Theoretical Relationships Between Facial Features and CVP

Facial Feature	Do (+)	Wa (+)	Co (+)	Tr (+)	Ge (+)	Ag (–)	At (+)	Theoretical Prediction
1. Facial width-to-height	(+)			(–)	(+)	(+)		+2, –2
2. Sexual dimorphism	(+)	(+)	(+)	(–)				+2 / +1, –1
3. Averageness							(+)	+1
4. Symmetry			(+)	(+)			(+)	+3
5. (Dark) Color	(+)					(+)		+1, –1

6. Babyfaceness	(–)	(+)		(+)	(–)		+3, –1
7. High cheekbones			(+)				+1
8. Large eyes	(–)	(+)		(+)		(+)	+3, –1
9. Thin jaw	(–)				(–)		+1, –1
10. Mouth-chin distance	(+)		(+)			(–)	+2, –1
11. Mouth-nose distance			(–)		(+)	(–)	–3

Notes. The personality traits we consider are Do: dominance, Wa: warmth, Co: competence, Tr: trustworthiness, Ge: generosity, Ag: aggressiveness, At: attractiveness; (+) and (–) denote positive and negative theoretical effects, respectively; the symbols under each personality trait in column headings are the trait's correlations with CVP, while the symbols within each row are the effects of the facial feature on the personality trait. The "Theoretical Prediction" column sums the contributions of the facial feature's effects on personality traits. For instance, in the first row, Facial width-to-height has a positive association with dominance, a negative association with trustworthiness, a positive association with generosity, and a positive association with aggressiveness. Dominance, trustworthiness, and generosity have a positive association with CVP, whereas aggressiveness has a negative association with CVP. Therefore, the overall impact of Facial width-to-height on CVP can be summarized as two positive forces (denoted as "+2"), from dominance and generosity, and two negative forces (denoted as "-2"), from trustworthiness and aggressiveness. Note that sexual dimorphism contains the effects of both masculinity and femininity; masculinity boosts dominance and competence for men, while femininity boosts warmth for women, so the overall theoretical prediction is "+2" for men and "+1" for women.

Feature 1: facial width-to-height ratio. This feature is calculated by dividing the face's width (bizygomatic) by its height (measured from the top of the eyelids to the upper lip). Previous studies have shown that the width-to-height ratio has a positive effect on inferred aggressiveness (Carré, McCormick, and Mondloch 2009), a negative effect on inferred trustworthiness (Stirrat and Perrett 2010), a positive effect on inferred generosity (in settings of cooperation and advising: Stirrat and Perrett 2012), and a positive effect on inferred dominance (in an organizational setting: Wong, Ormiston, and Haselhuhn 2011). The width-to-height ratio should contribute positively to CVP if the effects of inferred dominance and generosity outweigh the effects of inferred trustworthiness and aggressiveness.

Feature 2: sexual dimorphism. This feature captures the extent to which the face is distinguishably masculine or feminine as opposed to androgynous. For male faces, masculinity has positive effects on inferred dominance and competence (Penton-Voak et al. 1999) but may indicate less warmth and less trustworthiness (Gangestad et al. 2004; Penton-Voak et al. 1999; Perrett et al. 1998). For female faces, femininity is inferred as less trustworthy by males seeking

a mate (Little et al. 2014). For both male and female faces, people associate femininity with warmth and more concern for others, and masculinity with competence and self-assertion (Gao, Mittal, and Zhang 2020; Wen et al. 2020; Zhang, Feick, and Mittal 2014). For males, if inferred dominance and competence have more influence than inferred trustworthiness, then sexual dimorphism (masculinity, not androgyny) should contribute positively to CVP. For females, if inferred warmth has more influence than inferred trustworthiness, then sexual dimorphism (femininity, not androgyny) should contribute positively to CVP.

Feature 3: averageness. Averageness is the extent to which the face's features align with the average features of all people of the same gender, race, and age. Average faces may be perceived as more attractive (Rhodes and Tremewan 1996), likely because evolutionary pressures favor characteristics close to the population mean (Langlois and Roggman 1990). If physical attractiveness has a positive correlation with CVP, we also anticipate a positive effect of averageness on CVP; however, we can only be agnostic regarding the effects of attractiveness per se or the effects of averageness above and beyond those of attractiveness.

Feature 4: symmetry. Facial symmetry is the visual similarity (in shape and color) between the left and right sides of the face. Symmetry has positive effects on inferred attractiveness (Alley and Cunningham 1991; Gangestad et al. 2004; Rhodes, Sumich, and Byatt 1999), competence (in social networking: Fink et al. 2005; Fink et al. 2006; Pound, Penton-Voak, and Brown 2007), and trustworthiness (Noor and Evans 2003). Attractiveness, inferred competence, and inferred trustworthiness are all expected to contribute positively to charisma, so we anticipate that facial symmetry will also have a positive correlation with CVP.

Feature 5: color. Color is the skin tone of the face. Darker skin is perceived as being associated with more dominance and higher status, especially for athletes and entertainers (Wade

and Bielitz 2005) but also with more aggressiveness (Eberhardt et al. 2006; Livingston and Pearce 2009). Thus, we predict that darker skin should contribute positively to CVP if its positive effect via inferred dominance outweighs its negative effect via inferred aggressiveness.

Feature 6: babyfaceness. Babyfaceness is the extent to which the face's features resemble a typical baby's features rather than a typical adult's. The defining characteristics are large eyes, a small nose, a high forehead, and a small chin. Babyfaceness has positive correlations with honesty and warmth (Gorn, Jiang, and Johar 2008) and negative correlations with inferred power, dominance, and aggressiveness (Livingston and Pearce 2009). Babyfaceness should have a positive correlation with CVP if its positive effects through warmth, trustworthiness, and aggressiveness outweigh its negative effects through dominance.

Feature 7: thin jaw. Jaw width is calculated as the distance between the two edges of the jaw. A broader jaw has positive correlations with inferred dominance and strength (Cunningham, Barbee, and Pike 1990) as well as aggressiveness (Třebický et al. 2013). A thin jaw should correlate negatively with CVP if its positive effect on inferred dominance outweighs its negative effect on inferred aggressiveness.

Feature 8: large eyes. Eye size is a relative measure of the size of the eye against the whole face. Large eyes are perceived as more attractive (specifically, the perception that one is "charming"; Alley and Cunningham 1991) and heighten inferences of warmth, trust, and submissiveness (the opposite of inferred dominance; Montepare and Zebrowitz 1998). If the effects of large eyes on warmth and trust, and on attractiveness outweigh the effects on dominance, then large eyes should correlate positively with CVP.

Feature 9: high cheekbones. The cheekbones, particularly the malar bones, support facial structure; a person has "high cheekbones" if their malar bones are located closer to the eyes. In

males, high cheekbones have positive effects on inferred competence and dominance (in social networking: Cunningham, Barbee, and Pike 1990). Females with low cheekbones are perceived as being less competent (in reproductivity and social networking: Cunningham 1986) than those with high cheekbones. Thus, we predict a positive correlation between cheekbone height and CVP.

Feature 10: mouth-nose distance. This is the vertical distance between the nose tip and the upper lip. Researchers have found that attractive faces usually have a shorter mouth-nose distance (Perrett, May, and Yoshikawa 1994) and that a longer mouth-nose distance may predict sarcasm (Tay 2014), which is a passive, verbal form of aggressiveness (Pickering, Thompson, and Filik 2018; Szymaniak and Kałowski 2020). Additionally, people with a shorter mouth-nose distance are perceived as being more focused and flawless, so they likely are perceived to be more competent (Dunn 2018). The effects of the mouth-nose distance on attractiveness, aggressiveness, and competence suggest a negative effect on CVP.

Feature 11: mouth-chin distance. This is the vertical distance between the bottom edge of the lower lip and the base of the chin. Although a shorter mouth-chin distance looks more attractive, a longer distance predicts higher inferred competence in financial affairs (Alley and Cunningham 1991; Perrett, May, and Yoshikawa 1994). Also, Sinko et al. (2018) showed that people with a shorter mouth-chin distance are perceived as being less dominant and more submissive. If the effects of the mouth-chin distance on inferred competence and dominance outweigh its effect on attractiveness, then mouth-chin distance should correlate positively with CVP.

While these literature streams support separate predictions for each of these facial characteristics considered in isolation, in reality, faces represent composites of differences in

each. If we wish to understand which features affect charisma, and which inferred personality traits in turn drive visual aspects of celebrity potential, we need to build a model that can assess multiple features at the same time. We next describe the model that we use to do this. First, we present our data construction process. Second, we detail our findings about the facial features that appear to be most robustly associated with a given face's likelihood of celebrity as captured in the CVP score.

CVP MODEL DEVELOPMENT

Data Construction

We used three datasets for celebrities and six datasets for noncelebrities. All are benchmark datasets in computer vision and face recognition research, and they contain photos of people who vary in pose, age, gender, and race. The noncelebrity face datasets contain photos from social media, the Chicago Research Laboratory (Ma, Correll, and Wittenbrink 2015), and the internet. The celebrity face datasets contain photos from daily life, movies, and other scenarios. Most faces are labeled with the subject's identity, date of birth, date of the photo, and a rich set of attributes such as facial landmark locations, attractiveness, lip size, nose size, and hair color (Liu et al. 2015). The datasets include celebrities from a wide range of industries, such as entertainment (actors and influencers), sports, politics, and business (Liu et al. 2015), though individual photographs are not labeled with the celebrity's occupation. To explore whether CVP varies by industry, we compared the average CVP of 1,000 images of actors from the IMDb dataset and 1,000 images of people with various other occupations that we obtained from the other datasets combined. We found no significant difference ($t(1998) = 1.48$, $p = .14$), suggesting there is no obvious systematic bias arising from the occupations of celebrities.

Next, we randomly sampled 22,000 face images (11,000 each of celebrities and noncelebrities), drawing a balanced sample of images from each original dataset (details are in Web Appendix A). Then, after shuffling the dataset, we selected 12,000 images for training (80%) and validation (20%), and reserved 10,000 images for testing. We used a subset instead of the full sample for computational efficiency. However, we ensured accuracy by incrementally expanding our training set size until the model achieved satisfactory performance on the validation set and testing set (95.92% accuracy); beyond that point, we found a diminishing return on accuracy with more than about 12,000 images, as shown in Figure 1. The testing accuracy on 20,000 images is indeed higher (98.30%), but the improvement is marginal given the magnitude of the increase in computation costs. Given this trade-off, we set the cutoff at the "elbow" of 12,000 (see Kaplan et al. 2020 for a discussion of diminishing returns on model performance).

Figure 1. Diminishing Returns on Model Performance with Increasing Data Size

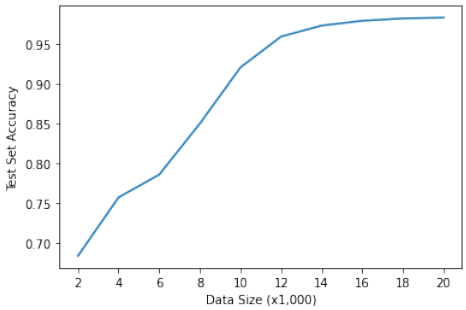


Table 2 presents the demographic composition of the datasets, with subjects categorized by celebrity status.[2] In Web Appendix A, we provide detailed descriptions of the datasets.

Table 2. Demographic Composition of Subjects in Datasets

Datasets	Gender		Age			
	Female	Male	< 20	20 – 40	40 – 60	> 60
Celebrity	29.43%	70.57%	.02%	82.28%	17.63%	.07%
Noncelebrity	28.62%	71.38%	.33%	93.56%	6.11%	.00%

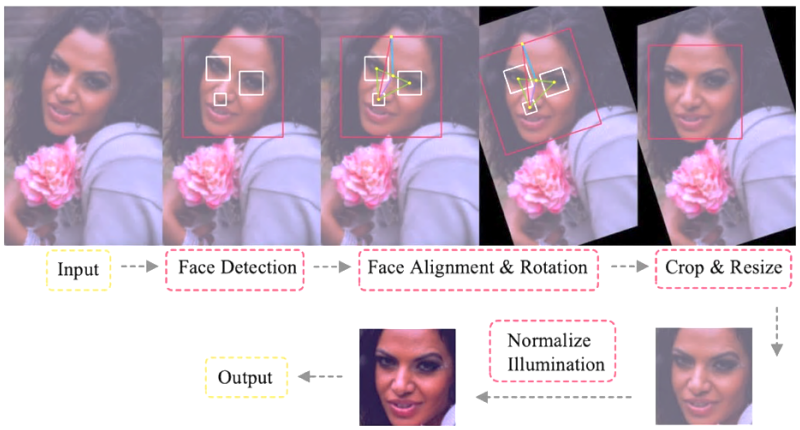
[2] The age, gender, and race of each face image was predicted using *Deepface*, which is based on the VGG-Face framework (Serengil and Ozpinar 2020).

Datasets	Race					
	Asian	Black	Indian	Latino / Hispanic	Middle Eastern	White
Celebrity	9.14%	7.27%	1.16%	9.36%	8.01%	65.06%
Noncelebrity	17.74%	5.98%	1.32%	11.55%	7.09%	56.33%

Data Preprocessing

We followed a five-step process to preprocess the images to address variations in size, face-background ratio, direction, and illumination (Figure 2). First, we detected the human face in the image. Second, we digitally aligned and straightened the face image (Kazemi and Sullivan 2014; Kovenko 2019) to reduce noise from variations in position and direction. Third, we cropped the image of each face to standardize the face-background ratio. Fourth, we resized the images to 224 × 224 × 3 pixels, the required input size for the pretrained ResNet-50 model. Last, we normalized illumination (Zhang, Lee, Singh, and Srinivasan 2021) to address systematic differences between datasets.

Figure 2. Face Image Preprocessing Steps



Notes. In preprocessing, we (1) detected the human face, nose, and eyes in the image; (2) rotated the image around the nose to straighten the face; (3) cropped the face from the image; (4) resized the image to 224 × 224 × 3 pixels; and (5) normalized illumination. Details of each preprocessing step and more examples are in Appendix B.

Model Development

We next developed our central construct, CVP, from the image data described in the previous subsection. To do so, we used a supervised deep learning model, shown in Figure 3 and described in detail in Web Appendix B. The input of the model is an image, and the output is the probability that the image is of a celebrity (i.e., CVP).

We employed ResNet-50 (He et al. 2016) as the backbone structure of the model; ResNet-50 is a computer vision deep learning architecture that was initially trained on more than a million images from the ImageNet database. We chose ResNet-50 for several advantageous features, including fast optimization and accuracy gain with model depth (i.e., number of layers). Since ResNet-50 was pretrained for facial recognition, as a related but distinct task, we adapted ResNet-50 to our analysis by adding two layers. One layer is a customized dense layer with 50 nodes, where one node is one neuron in a neural network model. We added this layer for model interpretation. The other layer is a sigmoid layer[3] that predicts CVP. We added the sigmoid layer to transform neuron values from real numbers to the range of 0 to 1, because CVP is defined as a probability. We further converted the continuously distributed CVP into a binary classification (i.e., celebrity or noncelebrity) using the threshold of .5, because our data are balanced across categories (Collell, Prelec, and Patil 2018).

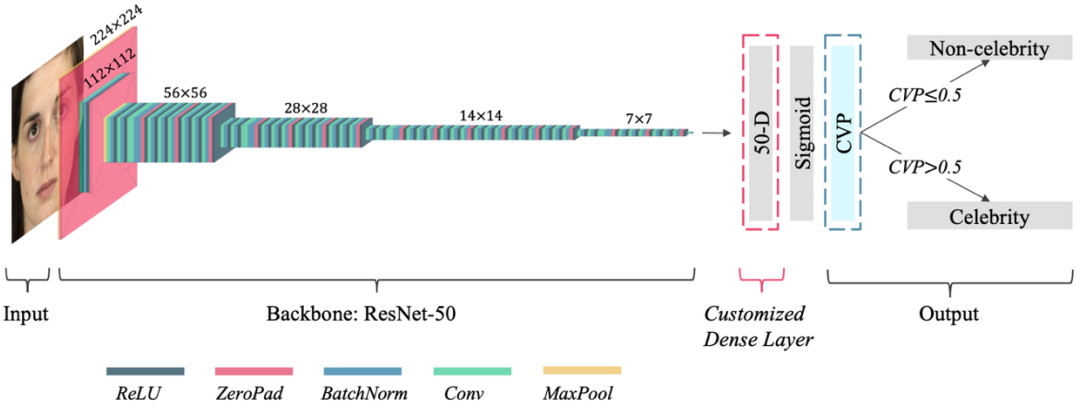
We also took several steps to optimize the model's performance. First, we selected the variant of the ResNet-50 model with the highest out-of-sample prediction accuracy and a stable optimization curve, or lower variation in convergence. Second, we experimented with five optimizers:[4] AdaGrad (Dean et al. 2012), SGD (Bottou 1998), AdaDelta (Zeiler 2012), Adam

[3] In the context of artificial neural networks, the term "sigmoid function" is an alias for the logistic function, which is calculated as $s(x) = 1/(1+e^{-x})$.

[4] An optimizer is an algorithm that reduces loss and improves accuracy by modifying the attributes of the neural network, such as weights and learning rate.

(Kingma and Ba 2015), and RMSprop. We achieved the highest accuracy and most stable optimization with the AdaGrad optimizer combined with preprocessing, as we explain in Web Appendix B.

Figure 3. Architecture of the Classification Model Based on ResNet-50

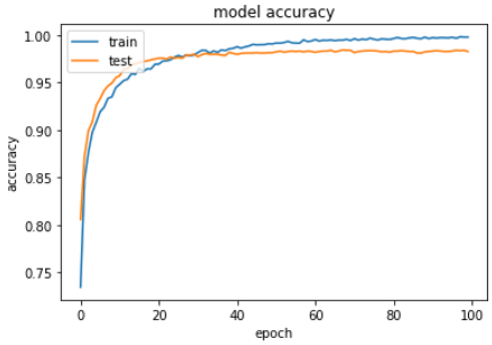


Notes. The structure of our classification model. The input is an image of 224 × 224 × 3 pixels. The backbone structure is ResNet-50, and the colors denote the layer types (see Web Appendix B). We added a customized dense layer with 50 nodes before the final sigmoid layer: the probability that the subject in the image is a celebrity (i.e., CVP). The 3D backbone was plotted using the Python package *visualkeras*.

Results

We used 12,000 images for training to calculate the model parameters, and reserved 10,000 for a holdout test set to assess the model's prediction accuracy. Based on results from the hyperparameter tuning stage (Web Appendix B), we achieved an accuracy of 95.92% on the test set. Figure 4 shows the improvement in accuracy on both the training and test sets during optimization. We also conducted a fivefold cross-validation to train five new models to make sure that our model has the optimal performance. A tenfold cross-validation on the optimized model demonstrates that our model has robust performance on different testing subsets (see details in Web Appendix B, pages 18-19).

Figure 4. Optimization Curves for the Training and Test Sets



Notes. The figure shows the optimization curves from the model with the best performance on the test set. We recorded the accuracy at each epoch. The blue line denotes the training accuracy, and the orange line denotes the accuracy on the hold-out set. The optimization curve became stable after epoch 40.

Table 3 presents summary statistics of CVP for each of the nine datasets (three celebrity datasets and six noncelebrity datasets) we used for training and testing.

Table 3. Summary Statistics on CVP for Different Datasets

Dataset	Type	Mean CVP	Max CVP	Min CVP
CelebA	Celebrity	.887	1.000	.002
IMDb-WIKI	Celebrity	1.000	1.000	.824
LFW	Celebrity	.993	1.000	.016
FEC	Noncelebrity	.214	1.000	.000
CFD	Noncelebrity	.001	.069	.000
GENKI-4K	Noncelebrity	.000	.003	.000
MTFL	Noncelebrity	.000	.002	.000
Selfie	Noncelebrity	.009	.983	.000
FFHQ	Noncelebrity	.003	.269	.000

Notes. Given the extraordinarily large size of each dataset, we calculated these summary statistics based only on the 22,000 images in our training and test sets. Dataset names are abbreviated; details are in Web Appendix A.

We benchmarked our model against two baseline models, support vector machine (SVM), and logistic regression. We tried two alternative inputs, including pixel-level vector and 11 facial features. For both types of input data, our model significantly outperforms both benchmarks on various performance metrics including prediction accuracy, recall, precision, and F1-score on the test set (details in Web Appendix B).

ASSOCIATION BETWEEN FACIAL FEATURES AND CVP

In this section we test the relationships between CVP and the 11 facial features discussed in the theoretical framework. First, we discuss the direction of the correlation between each facial feature and CVP, and then the contribution ranking of the facial features to CVP.

Direction of the Correlation Between Each Facial Feature and CVP

To test the relationships in Table 1, we created a contrasting dataset for each facial feature. That is, we selected or manipulated images to create two groups of images (Group 1 and Group 2) that varied in the focal feature. We then calculated the average CVP of the two groups using our deep learning model. For instance, for the facial width-to-height ratio, we digitally stretched the face in the image horizontally to create wider faces (Group 1) and vertically to create longer faces (group 2); the two groups had approximately the same facial features except for the width-to-height ratio. A higher average CVP in Group 1 (resp., Group 2) would indicate that at the population level, wider (resp., narrower) faces are associated with higher (resp., lower) CVP, implying a positive (resp., negative) correlation between the width-to-height ratio and CVP. We repeated this approach for all 11 facial features. The selection and manipulation processes and examples of contrasting images are provided in Web Appendix C, Table W2.

The group means and statistics appear in Table 4. CVP correlates positively and significantly with high cheekbones, (dark) color, large eyes, sexual dimorphism, and symmetry. CVP correlates negatively and significantly with facial width-to-height ratio and babyfaceness. However, there is no significant correlation between CVP and thin jaw, mouth-nose distance, mouth-chin distance, or averageness. Although CVP is predicted as a nonlinear function by the deep learning model of facial features, a dominant linear component may still exist in complex

and nonlinear relationships. If it exists, such a dominant linear component will be reflected in the population-level data.

Table 4. Direction of Correlation Between Each Facial Feature and CVP

Facial Feature	Correlation	Group 1	Group 2	t-Stat	*p*-Value
Facial width-to-height ratio	Negative	.953	.964	2.557	.01
High cheekbones	Positive	.014	.002	-63.039	.00
(Dark) Color	Positive	.006	.002	-36.803	.00
Thin jaw	Negative	.002	.009	1.544	.13
Mouth-nose distance	Negative	.002	.002	.240	.81
Large eyes	Positive	.005	.003	-2.387	.02
Sexual dimorphism	Positive	.633	.298	-19.187	.00
Mouth-chin distance	Positive	.002	.002	-1.578	.12
Babyfaceness	Negative	.001	.005	1.788	.08
Symmetry	Positive	.007	.001	-14.572	.00
Averageness	Negative	.897	.906	.959	.34

Notes. The table shows the direction of the correlation between each facial feature and CVP. The columns labelled "Group 1" and "Group 2" provide the mean CVP for the sets of selected or manipulated images explained in Web Appendix C. The "t-Stat" column compares the CVP distributions between Group 1 and Group 2 with a two-tailed test of the population mean with unknown variance. The "*p*-Value" column denotes the *p*-value for the corresponding t-statistics.

Ranking the Contributions of the Facial Features to CVP

To understand the relative importance of each facial feature to CVP, we leveraged SHAP (Lundberg and Lee 2017), a state-of-the-art explainable AI method. We went through a two-step process and used the customized dense layer with 50 nodes, as previously described.

In the first step, we used SHAP to calculate the contribution of each of the 50 nodes to CVP, namely NODESHAP. We assumed that the input node value from the second-to-last layer to the final sigmoid layer was the set of players, and the output was the payoff. We built a new XGBoost tree model (Chen and Guestrin 2016) and implemented the Tree SHAP algorithm. We chose Tree SHAP over Deep SHAP because the former is much faster in computation, computes the exact instead of approximated Shapley values (Shapley 1952), and allows for better visualization of the feature contribution (Lundberg et al. 2020). For additional details, see Web Appendix D.

In the second step, we calculated the contribution of each facial feature to CVP, denoted by FeatureSHAP. We derived FeatureSHAP based on the weighted NODESHAP, where the weights are the ranking of the node's relevance to the focal feature.

At the image level, the formula for $\text{FeatureSHAP}_i^{(p)}$ for feature i of individual p is

$$\text{FeatureSHAP}_i^{(p)} = \sum_{j=1}^{50}\{(51-j) \times \text{NODESHAP}_{n_{ijp}}\}, \quad (1)$$

where i denotes the i^{th} facial feature (= 1, 2, …,11); n_{ijp} (= 1, 2, …, 50) denotes the j^{th} (= 1, 2, …, 50) active[5] node (i.e., the j^{th} most relevant node to CVP) that captures the i^{th} feature for individual p; and $\text{NODESHAP}_{n_{ijp}}$ denotes the SHAP value for node n_{ijp} for individual p. For individual p, the weights (51– j) indicate that the most relevant node n_{i1p} (e.g., node 15) for feature i receives the highest weight, 50, whereas the least relevant node n_{i50p} (e.g., node 32) for feature i receives the lowest weight, 1.

At the population level, the FeatureSHAP for feature i is termed $\text{FeatureSHAP}_i^{(mean)}$:

$$\text{FeatureSHAP}_i^{(mean)} = \frac{1}{|P|}\sum_{p\in P}\sum_{j=1}^{50}\{(51-j) \times \text{NODESHAP}_{n_{ijp}}\}, \quad (2)$$

where P denotes the total population (i.e., all faces to be considered in the dataset), and |P| denotes the size of the total population (i.e., the total number of faces).

To determine the relevance of each node for each of the 11 facial features, we then used the following procedure. For each feature, we entered the contrasting groups of images (see Web Appendix C for details) and calculated the activation value of each node in the 50-D layer. Then, for each node, we calculated the Wasserstein distance[6] between the activation values for the

[5] In neural networks, whether a node is active is determined by its activation value—the higher the activation value, the more active the node is. This value is the output from a scalar-to-scalar activation function, which propagates the output of one layer's nodes to the next layer (Montesinos López, Montesinos López, and Crossa 2022).

[6] The Wasserstein distance is a distance function defined between probability distributions. For a definition, see Petitjean (2002) and Arjovsky, Chintala, and Bottou (2017).

contrasting groups (e.g., wider faces vs. longer faces). If a node captures the variation in this focal feature, then the node has a larger difference in activation values between the contrasting groups of images. Therefore, nodes with larger Wasserstein distances most strongly capture the predictive power of the focal feature (e.g., the facial width-to-height ratio) on the predicted CVP.

In Table 5, we rank the facial features (column 2) based on their FeatureSHAP values (column 3). Note that we focus on the predictive power of each facial feature here, not the level of significance in the correlation between each facial feature and CVP. We found that facial width-to-height ratio is the strongest predictor of CVP, followed by sexual dimorphism and averageness. This is because past literature has proved that significant variables are not always good predictors (Lo et al. 2015).

Table 5. Feature Importance Ranking Based on FeatureSHAP

Facial Features	Ranking	FeatureSHAP$^{(mean)}$
Facial width-to-height ratio	1	13.093
Sexual dimorphism	2	11.993
Averageness	3	10.408
High cheekbones	4	8.894
Color	5	8.494
Thin jaw	6	8.141
Mouth-chin distance	7	7.882
Large eyes	8	7.720
Symmetry	9	7.697
Mouth-nose distance	10	7.660
Babyfaceness	11	7.520

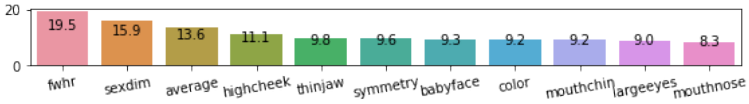
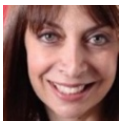
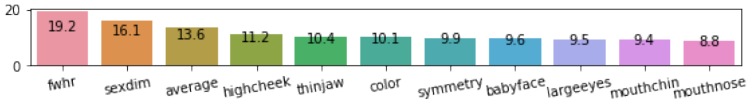
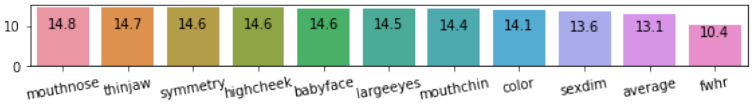
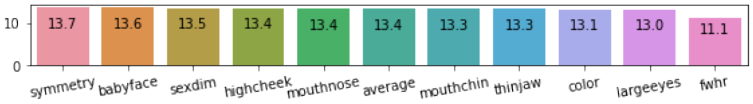
Significance levels: *$p < .1$, **$p < .05$, ***$p < .01$.

Individual heterogeneity. Previously, we have estimated the relationship between facial features and CVP at the population level, but individuals are heterogeneous in the contribution of each facial feature to CVP. For example, a thin jaw could be the most critical factor for one face and could have only a marginal influence on the CVP of another face.

Table 6 presents examples from these datasets, where the contribution of each feature to CVP is plotted but does not indicate the direction of these relationships. Combining this table

with Table 4, which presents the correlation between CVP and facial features at the population level, we derive interpretations such as the following: If Table 6 shows that an individual with a high CVP has Feature X (where X represents one of the 11 features) ranked highly, and Table 4 indicates a positive correlation of Feature X with CVP at the population level, we infer that this individual's high CVP is partly due to a high value of Feature X. For example, in the first two rows of Table 6, facial width-to-height ratio and sexual dimorphism are ranked highly. According to Table 4, facial width-to-height ratio correlates negatively with CVP, while sexual dimorphism correlates positively with CVP. Thus, individuals in these rows exhibit high CVP, attributed to their low facial width-to-height ratio and high sexual dimorphism. The CVP of the female in the last row is also driven down by her relatively small eyes; and the CVP of the male in the third row is driven down by his high degree of averageness.

Table 6. Four Examples of CVP Values

Image	CVP	FeatureSHAP$^{(p)}$
	.80	
	.81	
	.19	
	.22	

Notes. Variable FeatureSHAP$^{(p)}$ is calculated at the individual level and thus the ranking for each individual is different from the FeatureSHAP$^{(mean)}$ ranking. Here "fwhr" means facial width-to-height ratio, "sexdim" means sexual dimorphism, "highcheek" refers to high cheekbones, "mouthchin" denotes mouth-chin distance, and "mouthnose" denotes mouth-nose distance.

Comparison Between Empirical and Theoretical Correlations

We used the literature and our empirical evidence to derive the likely relationships between facial features and CVP. A symbolic summary is in Table 7. Some of our empirical results are consistent with the theoretical predictions, while others contradict them. For instance, our results show that symmetry is positively associated with CVP, and mouth-nose distance is negatively associated with CVP, reinforcing theoretical predictions. However, we find that averageness is negatively associated with CVP, in contrast with the theoretical prediction. Detailed explanations are presented in Web Appendix E.

Table 7. Symbolic Summary of the Theoretical and Empirical Comparison

Facial Feature	Theoretical Prediction	Empirical Result
1. Facial width-to-height	+2, –2	(–)
2. Sexual dimorphism	+2 / +1, –1	(+)
3. Averageness	+1	(–)
4. Symmetry	+3	(+)
5. (Dark) Color	+1, –1	(+)
6. Babyfaceness	+3, –1	(–)
7. High cheekbones	+1	(+)
8. Large eyes	+3, –1	(+)
9. Thin jaw	+1, –1	(–)
10. Mouth-chin distance	+2, –1	(+)
11. Mouth-nose distance	–3	(–)

Notes. The "Theoretical Prediction" column sums the contributions of the facial feature's effects on the personality traits (e.g., "+2, –1" means that two personality traits make positive connections between the facial feature and CVP, while one trait makes a negative connection). Note that sexual dimorphism contains the effects of both masculinity and femininity; masculinity boosts dominance and competence for men, while femininity boosts warmth for women, so the overall theoretical prediction is "+2" for men and "+1" for women (see Section *THEORETICAL FRAMEWORK,* Feature 2, for details). We include in the "Empirical Result" column from Table 4 to facilitate comparison.

INTERNAL VALIDITY: EXPERIMENTS 1 AND 2

In order to validate the causal relationships suggested in the development of the CVP measure, we next report one controlled experiment and, briefly, results from a replicate. We used them to test whether systematic variations in model-predicted CVP as presented in morphed faces result in

corresponding variations in the human judgment of CVP. We also measured participants' personality trait inferences and attractiveness perceptions, which we tested as simultaneous mediators in our model. These tests allow us to better understand whether CVP metrics may provide explanatory power that moves beyond the aforementioned ambiguous predictions associated with attractiveness. We describe Experiment 1 in detail and refer briefly to Experiment 2, which was a conceptual replication with a slightly different dependent measure. For the full details of Experiment 2, see Web Appendix F.

Experimental Design and Procedure

Participants. We recruited 1,153 US-based participants (43.06% male, 74.88% White) from Prolific.com. We excluded invalid, unfinished, or low-quality responses (see Web Appendix F for selection criteria), leaving 1,065 responses for analysis.

Stimuli. We used face-morphing (Venkatesh et al. 2021) to transform each of a set of faces (Black female, Black male, White male, White female) so their CVP values equaled 0, .2, .4, .6, and .8, yielding a total of 20 images (Table 8). The detailed procedure on face morphing is in Web Appendix H.

Table 8. Examples of Face Morphing to Achieve a Certain CVP Threshold

Face No.	Original Face	CVP Threshold	Morphed Face	CVP Threshold
1 Black Female		0		.6
2 Black Male		0		.6
3 White Female		0		.6
4 White Male		0		.6

Notes. The second column of the table shows the four original faces, whose CVP values are 0. The fourth column shows the modified faces after face-morphing transformations to the CVP levels of .6. Other examples of morphed faces with CVP values of .2, .4, and .8, are provided in Web Appendix F.

Procedure. Each participant was randomly assigned one morphed image from each of the four race-gender categories. For each of the four assigned faces, participants answered three questions about CVP as well as about measures of inferred personality traits. To measure CVP, we asked participants the following questions about each morphed image they were presented with: (1) How likely do you think that this person could become a celebrity? (2) If this person were on your favorite social media site, how likely is it that you would subscribe to their content or follow them? (3) This person is active on social media. Relative to other people, how many followers or subscribers do you think they have?

We used a 9-point Likert scale for responses, with higher numbers indicating a higher likelihood of having more followers. Participants also indicated their perceptions of the seven traits (i.e., dominance, warmth, trustworthiness, aggressiveness, competence, generosity, attractiveness) for each face on a 9-point Likert scale, with higher values indicating stronger inferred levels of each trait. Finally, participants provided demographic information about themselves including age, gender, race, state, political ideology,[7] social media usage, and financial scarcity.

Model

Using the observed variable OLS and logistic regression path analysis modeling tool, PROCESS Model 4 (5,000 bootstrapped samples; Hayes 2013), we explored the simultaneous parallel mediation effect of seven mediators—six personality traits (i.e., dominance, warmth, trustworthiness, aggressiveness, competence, generosity) and attractiveness—on the relationship

[7] The political ideology measure comprises 10 questions from a standard Pew Research Center survey, where option 0 indicates a conservative position and option 1 indicates a liberal position. The sum of the scores indicates the strength of the person's liberal position. See https://www.pewresearch.org/politics/2014/06/12/appendix-a-the-ideological-consistency-scale/.

in terms of how model-predicted CVP predicts human-rated CVP. We also controlled for the gender and race of the targeted face, as well as participants' gender, race, age, political ideology, social media usage, finance scarcity, and residence state.

Results

We combined participants' responses to the three CVP questions (Cronbach's alpha = .75) into a single index of human-rated CVP. The correlation between model-predicted CVP and human-rated CVP is .51 (p = .02). Figure 5, which shows the relationship between human-rated CVP and model-predicted CVP, demonstrates a positive and significant linear relationship (slope = .6699, p = .02) between the two. The figure shows model-free evidence that model-predicted CVP correlates positively with human-rated CVP. The y-axis represents the average human-rated CVP of each targeted face (about 210 ratings per face). The x-axis represents model-predicted CVP at five different levels. Each dot on the plot represents the average human-rated CVP for each of the 20 targeted faces. The blue line is the linear trend we obtained by fitting a regression model on these points with the formula Average Human-Rated CVP ~ Model-Predicted CVP.

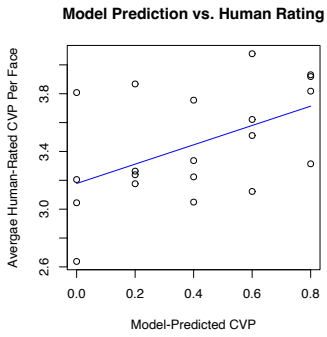


Figure 5. Scatterplot of Human-Rated CVP versus Model-Rated CVP

The results from Experiment 1 are shown in Table 9. The direct effect of model-predicted CVP on human-rated CVP is .05 ($F(17, 4218) = 236.00$, $p < .001$), and the total mediation effect

of the 7 mediators is .09. After accounting for the effect of attractiveness, we find significant indirect effects of model-predicted CVP through inferred dominance, warmth and competence and marginally significant mediation effects through inferred trustworthiness. Meanwhile we do not observe a significant mediation effect through inferred generosity or aggressiveness. Results are robust to the exclusion of covariates (Web Appendix F).

Table 9. Direct and Indirect Effect Analysis in Experiment 1 with Covariates

Variables	**Effect**	**SE (Bootstrapped)**	***p*-Value**	**95% CI (Bootstrapped)**
		Direct Effect		
Model-Predicted CVP	.047	.008	.000	[.031, .063]
		Indirect Effects		
Total	.086	.008	.000	[.071, .101]
Dominance	.014	.003	.000	[.008, .019]
Warmth	.011	.002	.000	[.007, .016]
Trustworthiness	.002	.001	.023	[.000, .004]
Aggressiveness	-.000	.001	.420	[-.002, .002]
Competence	.006	.001	.000	[.003, .009]
Generosity	.002	.001	.023	[-.000, .004]
Attractiveness	.052	.005	.000	[.043, .062]

Notes. The dependent variable is human-rated CVP. Covariates on demographic information of targeted faces and participants are included but not displayed. The p-values in column 3 are approximated based on Wald t-tests by dividing the estimate by the bootstrap standard error to arrive at a t-statistic (see practice in Abbua and Gopalakrishna 2021), but they are not accurate if the sampling distribution of the statistic is not symmetric. Thus, some papers only report bootstrapped CIs (e.g., Grewal, Hmurovic, Lamberton, & Reczek 2019). Furthermore, for concerns that faces of CVP = 0.8 seems to have makeup which might be a confounding factor, we conduct the analysis using faces with CVP < 0.8 and still find that the direct effect of model-predicted CVP on human-rated CVP is .04 ($p < .001$). Note that we did not "add" make up to faces, but rather it was part of the morphing process to change CVP of the same face. The detailed step-by-step tutorial on face morphing is in Web Appendix H. The selection of faces is random from the datasets and any attempt to select them on any specific dimension will not be appropriate. The detailed result is in Web Appendix G.

In Experiment 2, we used the same stimuli but showed the four faces side by side. Moreover, for the questions on CVP and personality traits, participants ranked the four faces instead of rating them individually. The results, in Web Appendix F, are consistent with those of Experiment 1.

As an alternative way of validation, we report average human-rated and model-predicted CVPs of faces ranked at 1, 2, 3, and 4 (Table 10). The average model-predicted CVP bears the same ranking order, showing that human judgment and model prediction are the same at the

population level; the average human-rated CVP from Experiment 1 bears the same ranking order, suggesting consistency between the two experiments.

Table 10. Average Human-Rated CVP and Model-Predicted CVP of Faces Ranked at 1, 2, 3, and 4 in the Lab Experiment

Human-Ranked CVP	Average Model-Predicted CVP	Average Human-Rated CVP
1	0.4846	3.7268
2	0.4340	3.5182
3	0.3712	3.3178
4	0.3116	3.2250

EXTERNAL VALIDITY: OBSERVATIONAL STUDIES 1 AND 2

While Experiments 1 and 2 suggest that general celebrity potential might be inferred from photographs, the nature of celebrity across domains may differ. For example, whereas greater Instagram "celebrity" would translate into higher post engagement (Tal and Gordon 2015), greater LinkedIn "celebrity" would be reflected in the likelihood of holding a C-suite position (Agle et al. 2006). Therefore, we undertook validation analyses on two external datasets that contained face photographs and outcomes that can be taken to represent domain-specific celebrities. We predicted that users with higher CVP would experience outcomes more consistent with domain-specific celebrities than those with lower CVP, even after we controlled for other contextual factors and measures of attractiveness.

Selfies of Instagram Influencers

To mitigate concerns that top-tier influencers may possess more resources for capturing higher-quality photographs compared to influencers with significantly fewer followers, we conducted data collection on the top 500 Instagram influencers of 2024. Within this cohort, only 230 influencers posted either a selfie or a photograph showing a clearly identifiable frontal face within our data collection period. Given that each of these influencers had more than one million users, disparities in image quality should not be a substantive concern for the validity of our

study. We extracted and preprocessed the face photographs as described in our initial machine-learning study. Then we used our deep learning model to predict the CVP of each face. We estimated the following model, where Follower # Change of a given influencer was tracked around one week after the time of posting the focal image, with a three-day gap (Cheng and Zhang 2024):

$$\text{Follower \# Change} = \text{Intercept} + \alpha_1 \times \text{Gender} + \alpha_2 \times \text{Age} + \alpha_3 \times \text{CVP} + \alpha_4 \times \text{Controls\_Contextual} + \alpha_5 \times \text{Facial\_Beauty} + \alpha_6 \times \text{Like \#} + \varepsilon. \quad (3)$$

We used the age-gender classification model proposed by Levi and Hassncer (2015) to measure Age (continuous number) and Gender (male = 1, female = 0). Here Controls_Contextual captured contextual data that might have correlated with the change in the number of followers: image visual features (Zhang, Lee, Singh, and Srinivasan 2021), including brightness, colorfulness, and symmetry, scored by the pyaesthetic package in Python;[8] the text description's length and richness (type-token ratio proposed by Chotlos 1944); and the text's sentiment, predicted using the compound score by VADER Sentiment Intensity Analyzer (Hutto and Gilbert 2014). We included physical attractiveness, measured by the ResNet-50 framework trained on the SCUT-FBP5500 dataset (Liang et al. 2018), to tease out the impact of attractiveness from the impact of CVP on Follower #. Additionally, we included the logarithm of like count (Like #) as a control variable. We estimated the full regression model as well as the models after feature selection through Step AIC (Hocking 1976). Table 11 presents the estimation results of the full regression model and the models after feature selection through Step AIC.[9]

[8] *Pyaesthetic* provides estimates of visual features concerning the aesthetic of a still image (https://github.com/Gabrock94/pyaesthetics).

[9] AIC refers to the Akaike information criterion; see Hocking (1976).

Table 11. Regression Results for the Instagram Dataset: Full Model and Step AIC Model

	Model (1)		Model (2)		Model (3)	
Variables	Full	Step AIC	Full	Step AIC	Full	Step AIC
CVP	120974.79 *	108678.10 .	113351.25 .	108678.10 .	9.00E+04	8.07E+04
	[60076.46]	[58987.41]	[59939.54]	[58987.41]	[55765.17]	[55062.47]
Attractiveness	-1.30E+04				-5.34E+03	
	[9588.71]				[9142.35]	
Gender (Female)	-2.98E+03		-3.47E+03		-9.90E+03	
	[16632.93]		[16664.09]		[16120.86]	
Age	-4.75E+03	-5.16E+03	-4.35E+03	-5.16E+03	-3.16E+03	
	[3808.30]	[3634.18]	[3804.71]	[3634.18]	[3558.28]	
Context Brightness	-7.21E+04		-6.13E+04			
	[69084.49]		[68774.04]			
Context Colorfulness	-592.80 .	-544.87 .	-556.03 .	-544.87 .		
	[323.48]	[317.44]	[323.03]	[317.44]		
Context Symmetry	1334.26 *	1347.26 *	1259.54 .	1347.26 *		
	[651.63]	[632.94]	[650.67]	[632.94]		
Description Length	-1.53E+01		-1.32E+01			
	[52.49]		[52.57]			
Description Richness	-5.22E+03		-1.63E+03			
	[25462.97]		[25378.73]			
Description Sentiment	3.73E+03		4.71E+0 3			
	[21814.51]		[21848.52]			
Like #	13109.71 *	13872.08 **	13339.72 *	13872.08 **		
	[5380.20]	[5097.27]	[5388.89]	[5097.27]		
R-Squared	9.46E-02	8.22E-02	8.62E-02	8.22E-02	1.61E-02	9.30E-03

Notes. The table presents the regression results of the full model and Step AIC model for the external Instagram dataset with and without contextual controls and variable Attractiveness. Standard errors are in brackets. Variables are standardized. Significance levels: *$p < .1$, **$p < .05$, ***$p < .01$.

CVP has a positive and significant coefficient in all cases, indicating that higher CVP predicts a higher increase in the number of followers, even after we controlled for variable Attractiveness and various features of post content (variable Controls_Contextual).

However, our results may be affected by selection bias, because people who choose to become influencers may generally have higher CVP than the average population. To verify

whether selection bias is a concern for our study, we randomly selected 1,000 photographs from one of the noncelebrity datasets (the Selfie Dataset, not included in the training set) and calculated the average CVP. Additionally, we randomly selected 2,105 Instagram photographs posted between 2016 and 2020 by about 500 influencers with more than one million followers each. These noncelebrities had an average CVP of .11, while the 2,105 influencers averaged .71, confirming selection bias exists. We also investigated the relationship between variables CVP and Attractiveness in this Instagram dataset of 2,105 selfies and found a relatively low correlation of .17 ($p < .001$). This finding demonstrates that CVP and attractiveness are essentially different; specifically, CVP adds extra explanatory power above and beyond attractiveness. Another caveat is that there might exist other factors that influence the number of followers. For example, networking with brands would be absorbed by the idiosyncratic error term.

LinkedIn Profile Data

We also validated our CVP model on LinkedIn profile images of C-level executives ("C-suite") as compared to non-C-suite employees. We selected 30 Fortune 500 companies (e.g., Amazon, Apple, Walmart) and collected the profile images of 5 C-suite and 5 non-C-suite employees for each company. Then we used a similar analysis process as for the Instagram influencers. We found that the mean CVP of the C-suite employees (.85) is significantly higher than the mean CVP of the non-C-suite employees (.20, $p < .001$ in a lower-tail test of the population mean).

DISCUSSION

Our research represents the first empirical attempt to characterize the relationships between CVP and facial features. We built a deep learning model that predicts a person's CVP based on a photograph of their face, and the model achieved high accuracy on the hold-out test sample. To

offer theoretical and practical implications, we used facial analytics on a large dataset to determine the direction and strength of the correlation between each of the 11 facial features we considered and CVP. We also derived a CVP formula (Equations 1 and 2) that uses SHAP to rank features by contribution to CVP. Furthermore, we experimentally showed that model-predicted CVP aligned closely with the human judgment of CVP in lab experiments. This alignment reveals the role of inferences related to different personality traits in explaining the relationship between facial features and CVP. Finally, we found that model-predicted CVP is predictive of desirable outcomes in the contexts of media and entertainment (Instagram) and business (LinkedIn).

Theoretically, our research relates closely to the literature on charisma. We consolidate prior theories on the correlations between facial features and charisma with personality traits as mediators. Our results support some of the relationships identified in prior work (e.g., CVP is hurt by a higher facial width-to-height ratio) and contradict others (e.g., the correlation between averageness and CVP is theoretically positive but empirically not significant in our data). Notably, we also found that while some proposed personality traits do explain the relationship between facial features and CVP, others seem to play a weaker mediational role. Specifically, it is more complex inferences, such as those related to generosity and aggressiveness, that fail to show robust mediation when considered in concert with other simpler but likely highly correlated inferences, such as warmth and dominance. Moreover, we empirically showed that CVP goes far beyond attractiveness. This finding calls for a qualification of widespread assumptions that emphasize the key role played by physical attractiveness to celebrity reported in previous studies.

Methodologically, while black box deep learning models are hard to interpret, we deploy the FeatureSHAP metric to provide interpretable features that lead to higher CVP. Although simpler methods, such as regression, also provide interpretable results, they are less flexible in capturing nonlinear relationships (some relationships cannot be easily captured by a simple functional form) and have much poorer prediction performance. In comparison, our deep learning model is purely data-driven and can capture very complicated forms of relationship with much higher prediction accuracy on the hold-out test sample.

MANAGERIAL IMPLICATIONS

Prior to outlining various potential managerial uses of the celebrity visual potential (CVP), it is imperative to stress that developments in machine learning raise ethical issues and concerns. Analogous to following the ethical guidelines on online data acquisition, organizations must exercise due diligence in obtaining explicit consent from individuals for the collection, disclosure, and intended utilization of facial data. Any managerial relevance discussion presupposes full compliance with such requirements.

A couple of noteworthy aspects of CVP merit careful consideration. First, in the external validation, while the impact of CVP is positive and significant, it explains only modest variance. Therefore, firms would benefit from using CVP as one of several factors, many of which are focused on skills or tangible characteristics. Second, the CVP score should theoretically generalize to professions where persuasiveness is an important job characteristic. This is likely the reason we were able to validate it for online professionals and managers. The CVP score is also a fairly widely generalizable characteristic, as persuasion is helpful in many roles. However, we note that occupations where persuasiveness is not critical to job performance may not relate well to the CVP

measure as designed. We indicate managerial insights that CVP offers in domains where visual factor is crucial.

First, CVP can be leveraged by employees in their career choices. Better knowledge of one's celebrity potential may help individuals decide whether to pursue a specific career. Second, CVP can be used as a control variable in field experiments or randomized controlled trials (RCTs) to improve persuasiveness or celebrity influence. This is analogous to the incorporation of genetic information in RCTs with objectives centered on enhancing academic outcomes (Lee et al. 2018). Given the predictive power of CVP on various outcomes, such use can generate nontrivial gains in statistical power for the RCT (Rietveld et al. 2013). Third, CVP can help design faces of AI-generated characters—such as figures in advertisements (Dave 2023), fashion models (Iovine 2023), or educators (Pataranutaporn et al. 2022)—making them more persuasive in interactions with potential customers in digital settings. Fourth, CVP can be used to assess if a company's human resource professionals excessively use facial features in their selection of salesforce. Selecting candidates based on a high CVP score is unlikely to reinforce existing biases based on demographic factors, as we find no relationship between CVP and race, gender, religion, or other typical bias factors. Rather, if we observe a positive link between CVP and sales success, opting for candidates with a high CVP may help HR staff focus on candidates with the highest likelihood of being strong performers.[10] However, we emphasize that CVP is only one of the many factors that affect employee performance and should never be used as the sole criterion in screening candidates.

The limitations of our work provide opportunities for future research. First, our study measures CVP as a static metric, while variations in the associations between facial features and

[10] The prediction accuracy of our model on the testing set (10,000 images) across different race, gender, and age groups is very high and does not vary significantly across different demographic groups.

CVP may exist over time. Second, several facial features have theoretical correlations with more than one personality trait and subsequently exert composite effects (i.e., with both positive and negative components) on CVP. While our empirical results provide clarity regarding the overall effect, the precise nature of the trade-offs between personality traits warrants further exploration. Third, utilizing the CVP metric comes with data privacy concerns, as with the use of consumer data. Thus, it may be necessary for the firm to ensure proper disclosure and obtain consent from consumers for the collection of data required to estimate CVP. Laws and regulations vary across countries, and adherence to these regulations is crucial for the successful implementation of CVP. Finally, the importance of CVP varies based on occupation, but our dataset lacks this information. Additional research including this occupation information would extend the applicability of our work.

REFERENCES

Abbua, Haroon R., and Pradeep Gopalakrishna (2021), "Synergistic effects of market orientation implementation and internalization on firm performance: Direct marketing service provider industry," *Journal of Business Research*, 125, 851-863,

Agle, Bradley R., Nandu J. Nagarajan, Jeffrey A. Sonnenfeld, and Dhinu Srinivasan (2006), "Does CEO Charisma Matter? An Empirical Analysis of the Relationships among Organizational Performance, Environmental Uncertainty, and Top Management Team Perceptions of CEO Charisma," *The Academy of Management Journal*, 49 (1), 161–74.

Alley, Thomas R., and Michael R. Cunningham (1991), "Averaged Faces Are Attractive, but Very Attractive Faces Are Not Average," *Psychological Science*, 2 (2), 123–25.

Alexander, J. C. (2010), "The celebrity icon," *Cultural Sociology*, 4 (3), 323–336.

Antonakis, John, Marika Fenley, and Sue Liechti (2012), "Learning Charisma," *Harvard Business Review*.

Arjovsky, Martin, Soumith Chintala, and Léon Bottou (2017), "Wasserstein Generative Adversarial Networks," in *Proceedings of the 34th International Conference on Machine Learning*, Doina Precup and Yee Whye Teh, eds. Proceedings of Machine Learning Research, 214–23.

Arthur, Michael B., Svetlana N. Khapova, and Celeste P.M. Wilderom (2005), "Career Success in a Boundaryless Career World," *Journal of Organizational Behavior*, 26 (2), 177–202.

Avolio, Bruce J., and Bernard M. Bass (1988), "Transformational Leadership, Charisma, and Beyond," in *Emerging Leadership Vistas*, James Gerald Hunt, B. Rajaram Baliga, H. Peter Dachler, and Chester A. Schriesheim, eds. Lexington, MA: Lexington Books, 29–49.

Baker, Michael J. and Gilbert A. Churchill, Jf. (1977), "The Impact of Physically Attractive Models on Advertising Evaluations," *Journal of Marketing Research*, 14 (November), 538-555.

Banker, Rajiv D., Hui Ding, Rong Huang, and Xiaorong Li (2023), "Market Reaction to CEOs' Dynamic Hemifacial Asymmetry of Expressions," *Management Science*.

Beck, Ulrich (2012), "Redefining the Sociological Project: The Cosmopolitan Challenge," *Sociology*, 46 (1), 7–12.

Berscheid, Ellen, and Elaine Hatfield (1978), *Interpersonal Attraction*, Reading, MA: Addison-Wesley Pub. Co.

Berscheid, Ellen and Elaine Walster (1974), "Physical Attractiveness," in *Advances in Experimental Social Psychology*, ed. Leonard Berkowitz, Vol. 7, New York: Academic Press.

Bottou, Léon (1998), "Online Learning and Stochastic Approximations," *On-line Learning in Neural Networks*, 17 (9), 142.

Buss, Arnold H., and Mark Perry (1992), "The Aggression Questionnaire," *Journal of Personality and Social Psychology*, 63 (3), 452–59.

Carré, Justin M., Cheryl M. McCormick, and Catherine J. Mondloch (2009), "Facial Structure Is a Reliable Cue of Aggressive Behavior," *Psychological Science*, 20 (10), 1194–98.

Cattell, Raymond B. (1950), *Personality: A Systematic Theoretical and Factual Study,* 1st ed. New York: McGraw-Hill.

Chaiken, Shelly (1979), "Communicator Physical Attractiveness and Persuasion," *Journal of Personality and Social Psychology*, 37 (August), 1387-1397.

Chase, Alexander (2016), *Unleash Your Charisma to Improve Your Social Skills and Create Long Lasting Relationships with Everyone You Meet*. Scotts Valley, CA: CreateSpace Independent Publishing Platform.

Chen, Tianqi, and Carlos Guestrin (2016), "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* New York: Association for Computing Machinery, 785–94.

Chen, Zeyang, Yu-Jane Liu, Juanjuan Meng, and Zeng Wang (2023), "What's in a Face? An Experiment on Facial Information and Loan-Approval Decision," *Management Science*, 69 (4), 2263–83.

Cheng, Mengjie, and Shunyuan Zhang (2024), "Reputation Burning: Analyzing the Impact of Brand Sponsorship on Social Influencers." *Management Science* (forthcoming). (Pre-published online October 18, 2024.)

Chotlos, John W. (1944), "IV. A Statistical and Comparative Analysis of Individual Written Language Samples.," *Psychological Monographs*, 56 (2), 75–111.

Collell, Guillem, Drazen Prelec, and Kaustubh R. Patil (2018), "A Simple Plug-in Gagging Ensemble Based on Threshold-Moving for Classifying Binary and Multiclass Imbalanced Data," *Neurocomputing*, 275, 330–40.

Conger, Jay A. and Rabindra N. Kanungo (1987), "Toward a Behavioral Theory of Charismatic Leadership in Organizational Settings," *The Academy of Management Review*, 12 (4), 637.

Costa, Paul, Robert R. McCrae, and Theodore M. Dembroski (1989), "Agreeableness versus Antagonism: Explication of a Potential Risk Factor for CHD," in *In Search of Coronary-prone Behavior: Beyond Type A*, Aron Wolfe Siegman and Theodore M. Dembroski, eds. Hillsdale, NJ: Erlbaum, 41–63.

Cunningham, Michael R. (1986), "Measuring the Physical in Physical Attractiveness: Quasi-Experiments on the Sociobiology of Female Facial Beauty," *Journal of Personality and Social Psychology*, 50, 925–35.

Cunningham, Michael R., A. Barbee, and C.L. Pike (1990), "What Do Women Want? Facialmetric Assessment of Multiple Motives in the Perception of Male Facial Physical Attractiveness," *Journal of Personality and Social Psychology*, 59 (1), 61–72.

Dave, Paresh (2023), "Google Will Soon Show You AI-Generated Ads," *WIRED*.

Dean, Jeffrey, Greg Corrado, Rajat Monga, Kai Chen, Matthieu Devin, Mark Mao, et al. (2012), "Large Scale Distributed Deep Networks," in *Advances in Neural Information Processing Systems*, P. Bartlett, F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, eds. Red Hook, NY: Curran Associates, Inc.

Dion, Karen, Ellen Berscheid, and Elaine Walster (1972), "What is Beautiful is good.," *Journal of Personality and Social Psychology*, 24 (3).

Dunn, Billie Schwab (2018), "What Does Your Face Say About You? Expert Reveals the Personality Traits of Four Celebrities Based on Their Looks - and the Secrets to 'Reading' the People in Your Life," *DailyMail* (July 19), https://www.dailymail.co.uk/femail/article-5961433/An-expert-revealed-faces-celebrities-really-say-them.html

Eberhardt, Jennifer L., Paul G. Davies, Valerie J. Purdie-Vaughns, and Sheri Lynn Johnson (2006), "Looking Deathworthy: Perceived Stereotypicality of Black Defendants Predicts Capital-Sentencing Outcomes," *Psychological Science*, 17 (5), 383–86.

Eder, Donna (1985), "The Cycle of Popularity: Interpersonal Relations Among Female Adolescents," *Sociology of Education*, 58 (3), 154–65.

Faerber, Stella J., Jürgen M. Kaufmann, Helmut Leder, Eva Maria Martin, and Stefan R. Schweinberger (2016). "The role of familiarity for representations in norm-based face space." *PloS one* 11(5): e0155380.

Fink, Bernhard, Nick Neave, John T. Manning, and Karl Grammer (2005), "Facial Symmetry and the 'Big-Five' Personality Factors," *Personality and Individual Differences*, 39 (3), 523–29.

Fink, Bernhard, Nick Neave, John T. Manning, and Karl Grammer (2006), "Facial Symmetry and Judgements of Attractiveness, Health and Personality," *Personality and Individual Differences*, 41 (3), 491–99.

Gangestad, Steven W., Jeffry A. Simpson, Alita J. Cousins, Christine E. Garver-Apgar, and P. Niels Christensen (2004), "Women's Preferences for Male Behavioral Displays Change Across the Menstrual Cycle," *Psychological Science*, 15 (3), 203–7.

Gao, Huachao, Vikas Mittal, and Yinlong Zhang (2020), "The Differential Effect of Local–Global Identity Among Males and Females: The Case of Price Sensitivity," *Journal of Marketing Research*, 57 (1), 173–91.

Goffman, Erving (1959), "The Moral Career of the Mental Patient," *Psychiatry*, 22 (2), 123–42.

Gorn, Gerald J., Yuwei Jiang, and Gita Venkataramani Johar (2008), "Babyfaces, Trait Inferences, and Company Evaluations in a Public Relations Crisis," *Journal of Consumer Research*, 35 (1), 36–49.

Gray, Kurt, Adrian F. Ward, and Michael I. Norton (2014), "Paying it Forward: Generalized Reciprocity and the Limits of Generosity.," *Journal of Experimental Psychology: General*, 143 (1), 247–54.

Grewal, Lauren, Jillian Hmurovic, Cait Lamberton, and Rebecca Walker Reczek (2019), "The Self-Perception Connection: Why Consumers Devalue Unattractive Produce," *Journal of Marketing*, 83 (1), 89–107.

Grewal, Rajdeep, Sachin Gupta, and Rebecca Hamilton (2021), "Marketing Insights from Multimedia Data: Text, Image, Audio, and Video," *Journal of Marketing Research*, 58 (6), 1025–33.

Griffin, Angela M., and Judith H. Langlois (2006), "Stereotype Directionality and Attractiveness Stereotyping: Is Beauty Good or is Ugly Bad?" *Social Cognition*, 24 (2), 187–206.

Hartmann, Jochen, Mark Heitmann, Christina Schamp, and Oded Netzer (2021), "The Power of Brand Selfies," *Journal of Marketing Research,* 58 (6), 1159–77.

Hayes, A. F. (2013), "Introduction to mediation, moderation, and conditional process analysis: A regression-based approach," *Guilford Press*.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2016), "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–78.

Hocking, R.R. (1976), "A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression," *Biometrics*, 32 (1), 1–49.

House, Robert J., and Jane M. Howell (1992), "Personality and Charismatic Leadership," *The Leadership Quarterly*, 3 (2), 81–108.

House, Robert J., William D. Spangler, and James Woycke (1991), "Personality and Charisma in the U.S. Presidency: A Psychological Theory of Leader Effectiveness," *Administrative Science Quarterly*, 36 (3), 364–96.

Hu, Allen, and Song Ma (2021), "Persuading Investors: A Video-Based Study," National Bureau of Economics Research (NBER) Working Paper No. 29048.

Hutto, C., and Eric Gilbert (2014), "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," *Proceedings of the International AAAI Conference on Web and Social Media*, 8 (1), 216–25.

Hwang, Serim, Xiao Liu, and Kannan Srinivasan (2021), "Voice Analytics of Online Influencers—Soft Selling in Branded Videos," *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.3773825

Iovine, Anna (2023), "Levi's to use AI models to 'increase diversity and sustainability,'" *Mashable*.

Kachur, Alexander, Evgeny Osin, Denis Davydov, Konstantin Shutilov, and Alexey Novokshonov (2020), "Assessing the Big Five Personality Traits Using Real-Life Static Facial Images," *Scientific Reports*, 10 (1), 8487.

Kahle, Lynn R., Pamela M. Homer (1985), "Physical Attractiveness of the Celebrity Endorser: A Social Adaptation Perspective," *Journal of Consumer Research*, 11(4), 954–961.

Kahle, Lynn R., Richard A. Kulka, and David M. Klingel (1980), "Low Adolescent Self-Esteem Leads to Multiple Interpersonal Problems: A Test of Social Adaptation Theory," *Journal of Personality and Social Psychology*, 39 (September), 496-502.

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, et al. S (2020), "Scaling Laws for Neural Language Models," *ArXiv*, https://arxiv.org/abs/2001.08361

Kazemi, Vahid, and Josephine Sullivan (2014), "One Millisecond Face Alignment with an Ensemble of Regression Trees," in *IEEE Conference on Computer Vision & Pattern Recognition*, 1867–74.

Keating, Caroline F. (2002), "Charismatic Faces: Social Status Cues Put Face Appeal in Context," in *Facial attractiveness: Evolutionary, cognitive, and social perspectives*, Gillian Rhodes and Leslie A. Zebrowitz, eds. Westport, CT: Ablex Publishing, 153–92.

Keating, Caroline F. (2011), "Channelling Charisma through Face and Body Status Cues," in *Social Psychological Dynamics,* Derek Chadee and Aleksandra Kostic, eds. Kingston, Jamaica: University of the West Indies Press, 93–111.

Kingma, Diederik P., and Jimmy Ba (2015), "Adam: A Method for Stochastic Optimization," *ArXiv*, https://arxiv.org/abs/1412.6980

Kovenko, Volodymyr (2019), "How to Precisely Align Face in Python with OpenCv and Dlib," *Medium* (August 11), https://medium.com/@dsfellow/precise-face-alignment-with-opencv-dlib-e6c8acead262

Langlois, Judith H., and Lori A. Roggman (1990), "Attractive Faces Are Only Average," *Psychological Science*, 1 (2), 115–21.

Levi, Gil, and Tal Hassncer (2015), "Age and Gender Classification Using Convolutional Neural Networks," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 34–42.

Lee, James J., Robbee Wedow, Aysu Okbay, et al. (2018), "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals," *Nature Genetics*, 50 (8), 1112–21.

Li, Yiyi, and Ying Xie (2019), "Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement," *Journal of Marketing Research*, 57 (1), 1–19.

Liang, Lingyu, Luojun Lin, Lianwen Jin, Duorui Xie, and M. Li (2018), "SCUT-FBP5500: A Diverse Benchmark Dataset for Multi-Paradigm Facial Beauty Prediction," *2018 24th International Conference on Pattern Recognition (ICPR)*, 1598–1603.

Little, Anthony C., Benedict C. Jones, David R. Feinberg, and David I. Perrett (2014), "Men's Strategic Preferences for Femininity in Female Faces," *British Journal of Psychology*, 105 (3), 364–81.

Liu, Ziwei, Ping Luo, Xiaogang Wang, and Xiaoou Tang (2015), "Deep Learning Face Attributes in the Wild," *2015 IEEE International Conference on Computer Vision (ICCV)*, 3730–38.

Livingston, Robert W., and Nicholas A. Pearce (2009), "The Teddy-Bear Effect: Does Having a Baby Face Benefit Black Chief Executive Officers?" *Psychological Science*, 20 (10), 1229–36.

Lo, Adeline, Herman Chernoff, Tian Zheng, and Shaw-Hwa Lo (2015), "Why significant variables aren't automatically good predictors," *Proceedings of the National Academy of Sciences*, 112 (45), 13892–97.

Lundberg, Scott M., and Su-In Lee (2017), "A Unified Approach to Interpreting Model Predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Ulrike von Luxburg, Isabelle Guyon, Samy Bengio, Hanna Wallach, and Rob Fergus, eds. Red Hook, NY: Curran Associates, Inc., 4768–77.

Lundberg, Scott M., Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, et al. (2020), "From Local Explanations to Global Understanding with Explainable AI for Trees," *Nature Machine Intelligence*, 2 (1), 56–67.

Ma, Debbie S., Joshua Correll, and B. Wittenbrink (2015), "The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data," *Behavior Research Methods*, 47, 1122–35.

Marshall, D. P. (1997), *Celebrity and power: Fame in contemporary culture*. Minneapolis: University of Minnesota Press.

Montepare, Joann M., and Leslie A. Zebrowitz (1998), "Person Perception Comes of Age: The Salience and Significance of Age in Social Judgments," *Advances in Experimental Social Psychology*, 30 (C), 93–161.

Montesinos López, O.A., A. Montesinos López, and J. Crossa (2022), "Fundamentals of Artificial Neural Networks and Deep Learning," In: *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer, Cham.

Moraes, Marcela, John Gountas, Sandra Gountas, and Piyush Sharma (2019), "Celebrity influences on consumer decision making: new insights and research directions," *Journal of Marketing Management*, 35(13-14), 1159-1192.

Noor, Fahim, and David C. Evans (2003), "The Effect of Facial Symmetry on Perceptions of Personality and Attractiveness," *Journal of Research in Personality*, 37 (4), 339–47.

Pataranutaporn, Pat, Joanne Leong, Valdemar Danry, Alyssa P. Lawson, Pattie Maes, and Misha Sra (2022), "AI-Generated Virtual Instructors Based on Liked or Admired People Can Improve Motivation and Foster Positive Emotions for Learning," in *2022 IEEE Frontiers in Education Conference (FIE)*, IEEE, 1–9.

Peng, Lin, Siew Hong Teoh, Yakun Wang, and Jiawen Yan (2022), "Face Value: Trait Inference, Performance Characteristics, and Market Outcomes for Financial Analysts," *Journal of Accounting Research,* 60 (2), 653–705.

Penton-Voak, I.S., D.I. Perrett, D.L. Castles, T. Kobayashi, D.M. Burt, L.K. Murray, et al. (1999), "Menstrual Cycle Alters Face Preference," *Nature*, 399 (6738), 741–42.

Perrett, D.I., K.A. May, and S. Yoshikawa (1994), "Facial Shape and Judgements of Female Attractiveness," *Nature*, 368 (6468), 239–42.

Perrett, D.I., K.J. Lee, I. Penton-Voak, D. Rowland, S. Yoshikawa, D.M. Burt, et al. (1998), "Effects of Sexual Dimorphism on Facial attractiveness," *Nature*, 394 (6696), 884–87.

Petitjean, Michel (2002), "Chiral Mixtures," *Journal of Mathematical Physics*, 43 (8), 4147–57.

Pickering, Bethany, Dominic Thompson, and Ruth Filik (2018), "Examining the Emotional Impact of Sarcasm Using a Virtual Environment," *Metaphor and Symbol*, 33 (3), 185–97.

Potts, John (2009), "The Age of Media: Charisma and Celebrity," in *A History of Charisma*, London: Palgrave Macmillan UK, 159–81.

Pound, Nicholas, Ian S. Penton-Voak, and William M. Brown (2007), "Facial Symmetry is Positively Associated with Self-Reported Extraversion," *Personality and Individual Differences*, 43 (6), 1572–82.

Rein, Irving, Philip Kotler, Michael Hamlin, and Martin Stoller (2006), *High Visibility: Transforming Your Personal and Professional Brand*, 3rd ed. New York: McGraw Hill.

Rhodes, Gillian, Alex Sumich, and Graham Byatt (1999), "Are Average Facial Configurations Attractive Only Because of Their Symmetry?" *Psychological Science*, 10 (1), 52–58.

Rhodes, Gillian, and Tanya Tremewan (1996), "Averageness, Exaggeration, and Facial Attractiveness," *Psychological Science*, 7 (2), 105–10.

Rietveld, Cornelius A., Sarah E. Medland, Jaime Derringer, et al. (2013), "GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment," *Science*, 340 (6139), 1467–71.

Riggio, Ronald E. (1987), *The Charisma Quotient*: *What It Is, How to Get It, How to Use It.* New York: Dodd, Mead & Co.

Rojek, C. (2001), *Celebrity*. London, England: Reaktion Books.

Schulz, Marc-Andre, B.T. Thomas Yeo, Joshua T. Vogelstein, Janaina Mourao-Miranada, Jakob N. Kather, Konrad Kording, et al. (2020), "Different Scaling of Linear Models and Deep Learning in UKBiobank Brain Images versus Machine-Learning Datasets," *Nature Communications*, 11, 4238.

Serengil, Sefik Ilkin, and A. Ozpinar (2020), "LightFace: A Hybrid Deep Face Recognition Framework," in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 1–5.

Shapley, Lloyd S. (1952), *A Value for N-Person Games*, Santa Monica, CA: RAND Corporation.

Sinko, Klaus, Reinhold Jagsch, Claudio Drog, Wilhelm Mosgoeller, Arno Wutzl, Gabriele Millesi, et al. (2018), "Facial Esthetics and the Assignment of Personality Traits Before and After Orthognathic Surgery Rated on Video Clips," *PLoS One*, 13 (2), e0191718.

Stirrat, Michael, and David I. Perrett (2010), "Valid Facial Cues to Cooperation and Trust: Male Facial Width and Trustworthiness," *Psychological Science*, 21 (3), 349–54.

Stirrat, Michael, and David I. Perrett (2012), "Face Structure Predicts Cooperation: Men with Wider Faces Are More Generous to Their In-Group When Out-Group Competition Is Salient," *Psychological Science*, 23 (7), 718–22.

Swami, Viren, and Adrian Furnham (2008), *The Psychology of Physical Attraction.* New York: Routledge/Taylor & Francis Group.

Szymaniak, Kinga, and Piotr Kałowski (2020), "Trait Anger and Sarcasm Use," *Personality and Individual Differences*, 154, 109662.

Tal, Diana, and Avishag Gordon (2015), "Charisma Research, Knowledge Growth and Disciplinary Shifts: A Holistic View," *Society*, 52 (4), 351–59.

Tay, Liz (2014), "7 Things Your Face Says About You," *Business Insider Australia* (January 14), https://www.businessinsider.com/seven-things-your-face-says-about-you-2014-1

Třebický, Vít, Jan Havlíček, S. Craig Roberts, Anthony C. Little, and Karel Kleisner (2013), "Perceived Aggressiveness Predicts Fighting Performance in Mixed-Martial-Arts Fighters," *Psychological Science*, 24 (9), 1664–72.

Troncoso, Isamar and Lan Luo (2023), "Look the Part? The Role of Profile Pictures in Online Labor Markets," *Marketing Science*, 42 (6), 1080–1100.

Venkatesh, Sushma, Raghavendra Ramachandra, Kiran Raja, and Christoph Busch (2021), "Face Morphing Attack Generation and Detection: A Comprehensive Survey," *IEEE Transactions on Technology and Society*, 2 (3), 128–45.

Verčič, Ana Tkalac, and Dejan Verčič (2011), "Generic Charisma – Conceptualization and Measurement," *Public Relations Review*, 37 (1), 12–19.

Wade, T. Joel, and Sara Bielitz (2005), "The Differential Effect of Skin Color on Attractiveness, Personality Evaluations, and Perceived Life Success of African Americans," *Journal of Black Psychology*, 31 (3), 215–36.

Wang, Yilun, and Michal Kosinski (2018), "Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation From Facial Images," *Journal of Personality and Social Psychology*, 114 (2).

Weber, Max, Alexander Morell Henderson, and Talcott Parsons (1947), *The Theory of Social and Economic Organization,* 1st American ed. New York: Oxford University Press.

Wen, Fangfang, Bin Zuo, Shuhan Ma, Yian Xu, John D Coley, and Yang Wang (2020), "Do We See Masculine Faces as Competent and Feminine Faces as Warm? Effects of Sexual Dimorphism on Facial Perception," *Evolutionary Psychology*, 18 (4), 1474704920980642.

Winterich, Karen Page, Vikas Mittal, and Karl Aquino (2013), "When Does Recognition Increase Charitable Behavior? Toward a Moral Identity-Based Model," *Journal of Marketing*, 77 (3), 121–34.

Wong, Elaine M., Margaret E. Ormiston, and Michael P. Haselhuhn (2011), "A Face Only an Investor Could Love: CEOs' Facial Structure Predicts Their Firms' Financial Performance," *Psychological Science*, 22 (12), 1478–83.

Zeiler, Matthew (2012), "ADADELTA: An Adaptive Learning Rate Method," *ArXiv*, https://doi.org/10.48550/arXiv.1212.5701

Zhang, Mengxia, and Lan Luo (2022), "Can Consumer-Posted Photos Serve as a Leading Indicator of Restaurant Survival? Evidence from Yelp," *Management Science* (published online April 4), https://doi.org/10.1287/mnsc.2022.4359

Zhang, Shunyuan, Dokyun Lee, Param Vir Singh, and Kannan Srinivasan (2021), "What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features," *Management Science* (published online December 21), https://doi.org/10.1287/mnsc.2021.4175

Zhang, Shunyuan, Nitin Mehta, Param Vir Singh, and Kannan Srinivasan (2021), "Can an AI Algorithm Mitigate Racial Economic Inequality? An Analysis in the Context of Airbnb," *Marketing Science*, 40 (5), 813-820.

Zhang, Yinlong, Lawrence Feick, and Vikas Mittal (2014), "How Males and Females Differ in Their Likelihood of Transmitting Negative Word of Mouth," *Journal of Consumer Research*, 40 (6), 1097–1108.

Zhou, Mi, George H. Chen, Pedro Ferreira, and Michael D. Smith (2021), "Consumer Behavior in the Online Classroom: Using Video Analytics and Machine Learning to Understand the Consumption of Video Courseware," *Journal of Marketing Research*, 58 (6), 1079–1100.

Zhou, Yinghui, Shasha Lu, and Min Ding (2020), "Contour-as-Face Framework: A Method to Preserve Privacy and Perception," *Journal of Marketing Research*, 57 (4), 617–39.

CHAPTER III: VISUAL UNIQUENESS IN PEER-TO-PEER MARKETPLACES: MACHINE LEARNING MODEL DEVELOPMENT, VALIDATION, AND APPLICATION

ABSTRACT

Peer-to-peer (P2P) marketplaces have seen exponential growth in recent years, featuring unique offerings from individual providers. However, scalable quantification of visual uniqueness and their impacts on platforms like Airbnb remains largely unexplored. We address this gap by developing, validating, and applying an unsupervised machine learning model to automatically extract uniqueness from images and quantify its impact on demand. We first construct a machine learning model, informed by cognitive psychology, to assess visual uniqueness in 481,747 property images, achieving high accuracy and interpretability. Next, we validate our model through three studies involving various participant populations and methods, confirming that the model's predictions of visual uniqueness align with human judgment. Finally, we apply this model to demand data of Airbnb properties in New York City spanning 13 months. We find an inverted U-shaped relationship between visual uniqueness and demand, with two significant moderation effects: properties with higher response rates or overall ratings benefit more from visual uniqueness. This research provides valuable insights for P2P platforms like Airbnb, highlighting the strategic use of visual uniqueness to enhance visual appeal and market performance. It also offers a new methodological roadmap for integrating psychological insights into the development and validation of unsupervised machine learning models.

Keywords: visual uniqueness, Airbnb, unsupervised contrastive learning, interpretable machine learning, image analytics, visual processing, peer-to-peer marketplace

INTRODUCTION

Incorporating unique features distinguishes a product or brand from competitors and enhances its competitive advantage (Kang and Na 2020). Individuals' innate need to differentiate themselves from others (Snyder and Fromkin 1977) drives the consumption of unique products, which not only bolsters personal and social identity (Tian, Bearden, and Hunter 2001), but also provides novel experiences (Tian, Bearden, and Hunter 2001). This consumer pursuit of uniqueness motivates companies to further enhance product differentiation.

The emergence of peer-to-peer (P2P) marketplaces, predicted to generate $335 billion in global revenue by 2025 (Industry Research Biz 2023), has revolutionized product differentiation. Rather than solely relying on companies or brands to create and communicate product uniqueness (Bloch 1995; Lauga, Ofek, and Katona 2022), P2P marketplaces involve individuals in the design, production, and distribution of offerings (Dellaert 2019), creating much vaster and more nuanced product differentiation and unique consumption experiences. Consider Airbnb as an example. Individual homeowners or hosts post pictures and other information about their homes on the platform to offer accommodations. Unlike hotel chains that share uniform designs and amenities, Airbnb properties vary widely in size, floor plan, design and décor, amenities, and more, making every offering different from the others to a certain extent. Instead of knowing what to expect (as with a branded hotel), Airbnb guests form perceptions of individual properties and their prospective lodging experience by browsing property images (Zhang et al. 2022) and information presented on property pages (Chung et al. 2022). Given the importance of product uniqueness, it is crucial to understand how consumers visually discern unique offerings and determine the optimal level of uniqueness. This understanding can guide individual providers to leverage the value of uniqueness effectively. However, the idiosyncrasy of the offerings

themselves and the varied presentation styles of individual suppliers pose a significant challenge for marketers in effectively measuring uniqueness in the P2P market, let alone quantifying its impact on market performance.

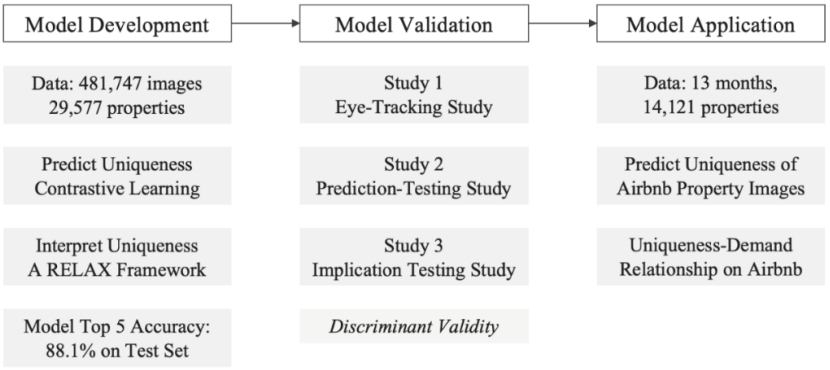
How do consumers process complex visual stimuli to discern uniqueness? Building on seminal works of visual processing (Berlyne 1970; Gibson and Gibson 1955; Itti and Koch 2001; Regehr and Brooks 1993; Theeuwes 2010), we conceptualize consumer perception of visual uniqueness, especially in a visually complex context where multiple idiosyncratic offers are presented, as a bottom-up, comparative, and holistic process without predefined standards. Operationalizing this process with computational contrastive learning (Chopra, Hadsell, and LeCun 2005), we configure an unsupervised machine learning model to analyze large-scale image data to measure visual uniqueness.

We optimize this model on 481,747 Airbnb property images. For every property image, the model automatically predicts a uniqueness score and generates a heat map that identifies the contribution of image features at the pixel level to visual uniqueness. The model's efficacy in mirroring human judgements is validated through three controlled studies with human participants, including an eye-tracking study. Finally, we apply the model to a dataset of 14,121 Airbnb properties in New York City (NYC) over 13 months (April 2022–April 2023) to examine the relationship between the uniqueness of property images and property demand, incorporating an extensive set of control variables. Additionally, we explore the moderating effects of host responsiveness and overall property rating, which can reduce perceived uncertainty in consumer choice—a significant concern, as increased uniqueness may heighten uncertainty perception (Bolton and Drew 1991; Ghose 2009).

This work lays out a roadmap for constructing and validating theory-grounded machine learning models, offers a scalable method for measuring visual uniqueness among a vast number of inherently diverse images, and provides actionable insights for operators in P2P marketplaces, where differentiation is both broad and nuanced.

The paper is structured as follows (see figure 1 for the paper workflow). We begin by delineating the process of visual uniqueness perception, drawing on interdisciplinary literature to establish a comprehensive understanding in visually complex settings, such as P2P marketplaces. Based on the conceptual process, the "Model Development" section outlines the construction, optimization, and interpretation methods of our unsupervised contrastive learning model, including a pixel-level analysis of the factors influencing image uniqueness scores. The "Model Validation" section focuses on validating the model's output against human judgment through three lab and online studies, including one with eye-tracking. We also compare our uniqueness score with image aesthetics and complexity metrics to establish its discriminant validity. In the "Model Application" section, we examine the relationship between our model-predicted visual uniqueness scores and consumer demand using Airbnb property demand data. Lastly, in the "General Discussion" section, we address potential concerns, limitations, and methodological and practical implications arising from our research.

FIGURE 1. PAPER WORKFLOW



FOUNDATIONS FOR MODELING VISUAL UNIQUENESS

The term "uniqueness" is used ubiquitously in popular and scholarly contexts. In marketing strategy, uniqueness is used synonymously with differentiation and distinctiveness in describing differentiated product or brand attributes (Carpenter, Glazer, and Nakamoto 1994). In consumer research, the concept of uniqueness is often discussed within the framework of "consumer need for uniqueness" through consumption (Tian, Bearden, and Hunter 2001), which has its roots in social psychology focusing on individuals' need to distinguish themselves from others (Snyder and Fromkin 1977). Additionally, related terms, such as "prototypicality" (Landwehr, Labroo, and Herrmann 2011), "originality" (Pieters, Warlop, and Wedel 2002), and "distinctiveness" (Cesareo, Townsend, and Pavlov 2023), have been studied in product design and advertising appeal contexts, where uniqueness is derived from deviation from category prototypes or judged by small samples of consumers. Most of these concepts imply that the perception of uniqueness results from comparative processes. In a visually complex environment like a P2P online marketplace, how do consumers process idiosyncratic visual stimuli, and what comparisons do they make in order to discern unique offerings? To model this process, we draw on a long tradition of cognitive visual processing research.

A Conceptual Model of Visual Uniqueness Perception

Bottom-up as the primary processing mechanism. Visual perception is generally understood as an interplay between bottom-up and top-down perceptual processes, with bottom-up processing taking precedence when individuals encounter novel stimuli (Theeuwes 2010). While top-down processing relies on pre-existing knowledge (Bouvier 2009), bottom-up processing involves the direct analysis of visual stimuli based on their inherent features, independent of prior knowledge or expectations (McMains and Kastner 2011). When consumers browse Airbnb or similar P2P platforms, they encounter idiosyncratic lodging offerings with diverse decorations, viewing angles, and object arrangements. Such a complex visual environment with novel stimuli makes it difficult for consumers to draw on pre-existing mental schemas for immediate judgments, thereby prompting bottom-up processing (Gibson 2002).

When viewing a single image, bottom-up processing starts with the perception and integration of basic visual elements such as color, contrast, texture, and shape (Itti and Koch 2001). When viewing multiple images, either on the same page or across different pages, this process evolves into a comparative evaluation of visual stimuli.

Comparison as the core evaluative process. The recognition of uniqueness inherently requires comparison. In visual perception, this entails identifying stimuli that stand out from the rest. Foundational theories in visual processing assume spontaneous comparisons among visual stimuli. Gibson and Gibson (1955) argued that differentiation is the fundamental process of visual processing, as individuals exposed to various stimuli learn to detect subtle differences and invariant features, enhancing perceptual efficiency and accuracy. Berlyne (1970) posited that a viewer's interest in a visual stimulus is not only influenced by its complexity, regularity, and congruity, but also by its repetition and sequence, emphasizing the critical role of these relative

visual characteristics, which he termed “collative variables.” This comparative evaluation process becomes even more pronounced when the image-viewing task itself is aimed at comparing different offerings.

Holistic individuation over feature-based analysis. How do individuals make comparisons when evaluating diverse P2P offerings? Regehr and Brooks (1993) proposed that when comparable stimuli display unique characteristics at the individual feature level, people tend to engage in holistic individuation. This involves recognizing and differentiating stimuli based on an overall perception of the stimuli as a whole, rather than analyzing and comparing individual features in isolation. They argued that when stimuli contain complex, idiosyncratic information, this complexity cannot be readily reduced to a few dimensions or abstract categories. Consequently, analytic processing based on predefined features may fail to capture the nuances critical for differentiation. Focusing exclusively on predefined features risks overlooking essential details that distinguish one stimulus from another, making holistic processing necessary to identify these idiosyncratic elements (Palmeri and Gauthier 2004).

Limitations of top-down and feature-based approaches. Past attempts to assess concepts related to uniqueness have primarily focused on deviations from top-down category norms or feature-based prototypes. For instance, Meyers-Levy and Tybout (1989) examined “schema incongruity” by analyzing how consumer responses vary when products deviate from existing category schemas—a top-down process. Similarly, Landwehr, Labroo, and Herrmann (2011) explored “prototypicality” and its associated “atypicality” by evaluating how products compare to category prototypes based on averages of predefined features. However, the idiosyncratic nature of P2P offerings defies such top-down approaches. The visual complexity and novelty

inherent in P2P markets challenge straightforward schema-based averaging, and the nuanced differentiation crucial to assess property uniqueness often escapes typical feature-based models.

Toward a bottom-up, comparative, and holistic model. In light of these limitations, we conceptualize the perception and evaluation of visual uniqueness as a process driven by bottom-up processing, comparative evaluation, and holistic individuation. This process assesses all visual features—such as color schemes, contrasts, object placements, and textures—not solely within individual images, as suggested by Itti and Koch (2001), but across images to highlight differences. Unlike traditional models that depend on preconceived schemas or predefined standardized features, such as those used by Landwehr et al. (2011), our approach focuses on the unique, emergent qualities of stimuli.

This conceptualization lays the groundwork for constructing our computational machine learning model, designed to assess visual uniqueness based on these principles.

Unsupervised Contrastive Learning: Unlocking the Potential for Identifying Visual Uniqueness

With advancements in machine learning, image analytics has become increasingly relevant in marketing research. However, the subject of image uniqueness remains comparatively understudied, a gap largely attributable to the absence of a labeled dataset necessary for model training. Even when human labels are available, they often present challenges due to their time-sensitive and context-dependent nature (Zhong et al. 2022). Furthermore, the complexity increases when interpreting what contributes to an image's uniqueness in a scalable manner.

To address these challenges, innovative algorithms and techniques are necessary to avoid reliance on human labeling and align with the conceptual model of visual uniqueness processing described earlier. One such approach is leveraging contrastive loss (Oord, Li, and Vinyals 2018), a foundational algorithm that guides unsupervised models (requiring no labeled data) in

distinguishing between similar and dissimilar samples, much like discerning apples from oranges in a basket of assorted fruits. It categorizes samples based on the "distance" (which can be measured using the cosine or Euclidean distance) between every pairwise sample and organizes them into clusters of similarity and dissimilarity in a high-dimensional representation space (Chopra, Hadsell, and LeCun 2005). This technique advanced image analysis, in particular, by capturing intricate attributes, such as color and size, and offering a more expressive, feature-rich representation that is capable of accommodating a wide array of model configurations and complexities (Le-Khac, Healy, and Smeaton 2020).

While contrastive learning has proven effective in diverse applications, such as face recognition (Schroff, Kalenichenko, and Philbin 2015) and scene classification (Sermanet et al. 2017), its specific utility for capturing visual uniqueness remains unexplored. Nevertheless, its potential to model perceptions of uniqueness is promising due to its intrinsic pairwise comparison approach. Unlike traditional top-down methods that compare new stimuli against established prototypes or exemplars on a fixed set of predefined features (e.g., Liu et al. 2017), contrastive learning allows stimuli to be compared with one another across all aspects, mimicking human's holistic bottom-up visual processing in between-stimuli comparisons.

Given the potential resemblance of contrastive learning to the process of visual uniqueness perception in visually complex P2P settings, we propose a machine learning model with the following three features: 1) A data-driven, unsupervised contrastive learning model that assesses full image information without predefined features, mimicking the bottom-up, comparative, and holistic process for uniqueness perception and evaluation. 2) Images are randomly drawn from subsets of a "memory bank" of the training set, mimicking human sampling of product images from browsing experience. 3) Unique characteristics of each image are interpreted at the pixel

level, rather than based on predefined features, considering all image information as contributing to visual uniqueness. The unsupervised approach eliminates the need for human labeling and enhances scalability, suitable for P2P marketplaces with large volume of offerings.

These three features set our work apart not only from the traditional prototype model, but also from prior attempts using machine learning to study image or product uniqueness (e.g., Cesareo et al. (2023)). In the next section, we describe the construction of our contrastive machine learning model and explain its features in detail.

MODEL DEVELOPMENT: QUANTIFYING VISUAL UNIQUENESS

We construct and optimize our computational model on a large image dataset from Airbnb, a leading P2P online platform facilitating short-term home rentals. As discussed above, we employ an unsupervised learning approach to circumvent issues with labeled data and the risk of distribution shift due to the dynamic nature of visual perception (Zhong et al. 2022). The model autonomously learns to discern and extract image features from unlabeled data, assigning a score to each image that quantifies its uniqueness within the dataset. Next, we outline the dataset, detail the model's construction process, and interpret the model prediction results.

Data Preparation. First, we identified Airbnb properties in NYC from 192 different zip codes active as of January 2022,[11] and collected their property images monthly from April 2022 to April 2023. This yielded a total of 481,747 property images, which were subsequently divided into a training set of 420,735 images and a holdout test set of 61,012 images. We first resized

[11] An active property is a property that had at least one day open in a month to be booked on Airbnb.

images to 224 × 224 pixels to ensure uniform size,[12] which is required by our model, and normalized their illumination to minimize the impact of varying light conditions.

Next, we prepared similar image pairs to replace human judgment data as "labels" for the original image data. We employed four random data augmentations to create variants of the same image as similar pairs: random color jitter (randomly changing the brightness, contrast, and saturation of an image), gray scale (turning an image into grayscale where the value of each pixel carries only intensity information), Gaussian blur (using a Gaussian function to calculate the transformation to apply to each pixel in an image), and horizontal flip (creating a horizontally mirrored version of the original image; Shorten and Khoshgoftaar 2019).[13] Using similar image pairs as labels in visual analytics provides an efficient and scalable way to train machine learning models to understand visual (dis)similarity—the basis of visual uniqueness. This approach enhances feature learning by focusing on relationships among images rather than fixed categories, enabling models to identify underlying patterns that define similarity in a bottom-up fashion, aligning with our conceptualization. It simplifies the labeling process, eliminating the need for costly human labor and providing robust generalizability across contexts and time.

Contrastive Learning Model Optimization. For computation efficiency, we first tuned hyper-parameters (e.g., learning rate and batch size) on a randomly drawn subset of the training data before optimizing the model on the full training set (see Web Appendix B). Next, on the full training data, we trained our unsupervised model to extract visual uniqueness from input images by comparing each image against others and computing a contrastive loss for each image.

[12] The cropping retains the largest central square of the image (Krizhevsky, Sutskever, and Hinton 2017). This method standardizes various image sizes and orientations while preserving most of the critical visual information (Lester 2013).
[13] The four augmentations were chosen for their ease of implementation and computational efficiency in diverse contexts and datasets (Poynton 1998; Qian et al. 2020). See Web Appendix B for technical details.

Figure 2 illustrates the construction and training procedures of the model. Based on our conceptualization of visual uniqueness perception as a bottom-up, comparison-based, holistic individualization process where people scan the entire images, compare images to their immediate surroundings to detect features that make each image stand out from others (Theeuwes 2010; McMains and Kastner 2011), we construct a machine learning model to mimic the above perception process. The machine learning model also implements a bottom-up, holistic comparison of images to determine the uniqueness of each image, achieved with the four steps.

Step 1: Creating Similar and Different Image Pairs. To train the model to recognize whether two images are similar or different, we generated the following pairs: 1. Similar Pairs: These were created by applying random augmentation (e.g., blurring, horizontal flipping) to the same image. These augmentations ensure the images look slightly different while still maintaining similarity (see [a_1] of figure 2). 2. Different Pairs: These were generated by pairing each image with unrelated images randomly drawn from the entire image dataset (see [a_2] of figure 2). This approach eliminates the need for human-labeled data, making our machine learning model unsupervised. It also ensures the framework is adaptable to any context without requiring pre-existing human judgment data.

Step 2: Extracting and Storing Vector Representation. For the two images in each pair, we use feature encoders to extract essential details from each version. These encoders reduce the image size by capturing only the most important information, converting each large image into a compact representation of 128-dimensional features. The feature extraction mimics bottom-up visual processing, where raw visual elements are identified and stored in short-term memory, ready for comparison (see [b] and [c] of figure 2).

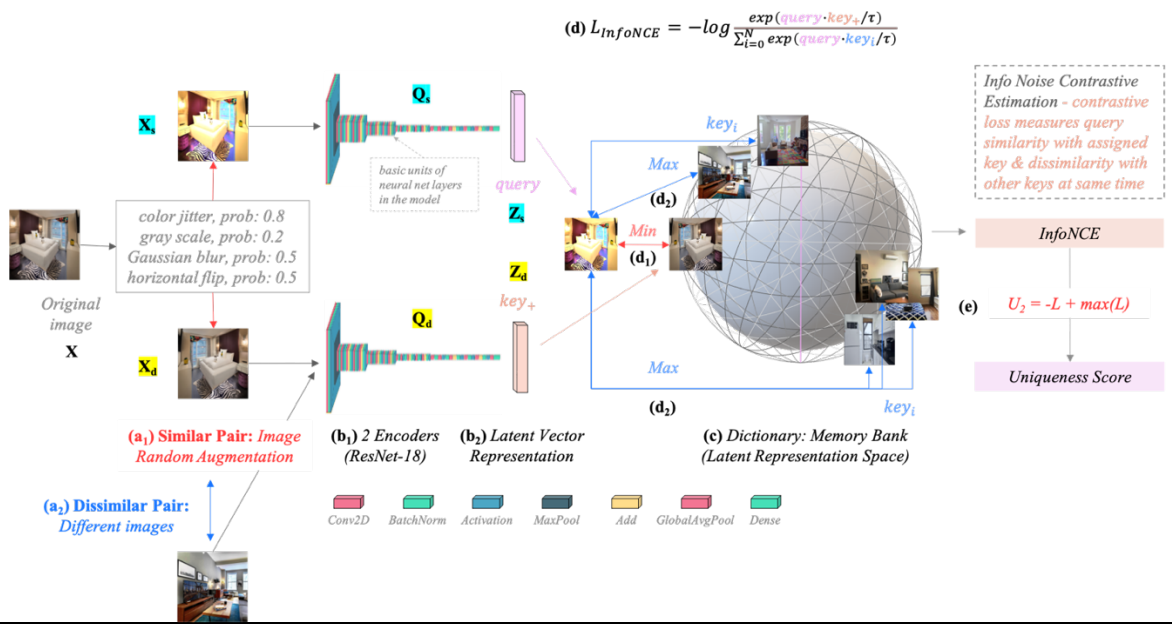
Step 3: Calculating Contrastive Loss. We use a metric called "InfoNCE" to measure how similar or different two images are. This calculation considers two aspects simultaneously: 1. Similarity: It evaluates how close two slightly modified versions of the same image are. Since these versions originate from the same image, they should remain relatively similar. 2. Dissimilarity: It examines how distinct one version of an image is compared to unrelated images in the dataset. The Similarity and Dissimilarity metrics are captured by parts (d_1) and (d_2), in equation (d) of figure 2, respectively. The model's objective is to minimize the InfoNCE value by: 1) "pulling" similar image pairs closer together so they are recognized as "similar," effectively aligning the vector representations of two similar pairs, and "pushing" dissimilar image pairs further apart to clearly distinguish them as "different." This process mirrors how humans might compare images, recognizing uniqueness by contrasting one image with another. The feature vector representation is visualized as points on a sphere, where similar images cluster together while dissimilar images spread apart. This process mimics how working memory rapidly organizes images based on similarity. The computed InfoNCE value, *L*, is the loss function of the machine learning model; the smaller the value, the more unique an image is.

Step 4: Uniqueness Score Transformation. We compute a "Uniqueness Score" for each image by adjusting the previously measured loss, *L*, using a linear transformation: Uniqueness = -*L* +intercept, where the intercept is the maximum value of loss among all images. The resulting score provides a clear and intuitive way to evaluate how much each image stands out in terms of visual uniqueness. This is explained in part [e] of Figure 2.

As discussed, the model learns uniqueness by contrasting similar and dissimilar image pairs. The model's learning is effective if the pairs constructed in Step 1 are reasonable: the two variants of the same focal image should be more similar to each other than to a randomly

selected image from the dataset. While random selection may introduce variability in dissimilar pairs, it's unlikely that a randomly chosen image would be more similar to the focal image than its augmented versions, except in cases of image duplication. However, this is not a concern for two reasons: First, image duplication is rare, as Airbnb actively detects and deactivates listings with duplicate images to prevent fraud. Second, our random sampling process ensures unbiased training, even in the case of duplicates. Each time, 256 images are randomly sampled from the dataset, generating over 8,000 samples per focal image. Any bias from duplicated images is effectively averaged out across these numerous samples.

FIGURE 2. STRUCTURE OF CONTRASTIVE LEARNING MODEL



NOTE: In (a), four data augmentation methods are implemented with specified probabilities and multiple augmentations that can have some probability of applying to the same image: color jittering is applied to an input image with an 80% likelihood; grayscale conversion with a 20% likelihood; Gaussian blurring with a 50% likelihood; and horizontal flipping with a 50% likelihood. The basic units of a neural network (e.g., Conv2D) are standard and not specific to our contrastive learning; therefore, they are not discussed (He et al. 2016).

Model Evaluation. Our model applies random data augmentation to create two variants of the same property image (e.g., one horizontally flipped, and the other Gaussian blurred). We follow standard contrastive learning procedures (Yin et al. 2023) and evaluate our model by its

ability to correctly pair these variants. High accuracy, defined as the probability of correct pairing, shows that the model effectively distinguishes between similar images (variants of the same original) and dissimilar images (from different originals). On the holdout test set of 61,012 Airbnb property images, the model achieves a top-1 accuracy of 73.1% (the model's top-1 or highest probability prediction is the original image),[14] which is comparable or superior to the state-of-the-art contrastive learning models (60% to 80%; Yin et al. 2023).

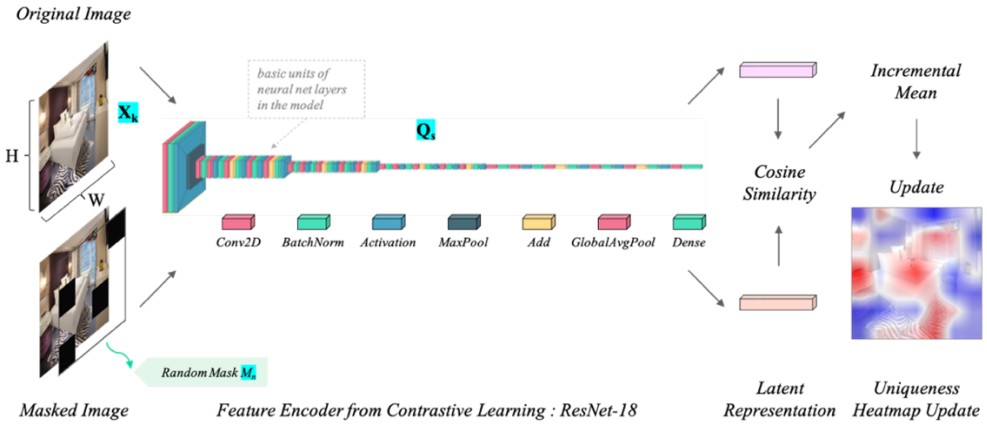
Model Interpretation. A common critique of machine learning models is their opacity, especially in image processing in which a complex, high-dimensional input is reduced to a single value, such as a predicted uniqueness score (Lipton 2018). This opacity often limits the practical value of these models, making it difficult to offer interpretable and actional advice for businesses to diagnose problems or develop strategies. To address this and move beyond treating the model's output as a "black box," we employ the *Representation LeArning eXplainability* (RELAX) framework (Wickstrom et al. 2021) to identify what makes an image unique (i.e., the drivers of uniqueness prediction) at the pixel level.

Figure 3 illustrates the RELAX framework. Consider $X_k \in R^{H \times W}$ as an image matrix defined by the pixel intensities in the RGB (red, green, blue) channels, with dimensions determined by the height H and width W. First, X_k is embedded into a 128-dimensional vector by the feature encoder Q_s. We then apply a set of random masks $\mathbf{M} = \{M_1, M_2, \ldots, M_n, \ldots\}$, where each M_n is a matrix of zeros and ones that represents the same size as the input image. We generate $M_n \in (0, 1)^{H \times W}$, with each element m_{kij} generated by a Bernoulli distribution with probability p ($m_{kij} \sim$ Bernoulli[p]). Multiplying image matrix X_k by M_n randomly "masks" pixels of image k—those multiplied by 0 become hidden. This alters the image's vector representation and its resulting

[14] The top-5 accuracy is 88.1%, where the original image is among the model's top five highest probability answers.

uniqueness prediction, enabling us to infer the contribution of specific pixels to the overall uniqueness score. By applying a diverse range of M_n, we can assess the impact of each pixel, creating a *uniqueness heat map* in which the significance of each pixel's contribution is indicated by color intensity at the pixel level. Web Appendix D presents the technical details.

FIGURE 3. STRUCTURE OF RELAX FRAMEWORK

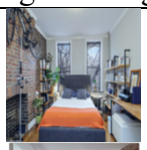
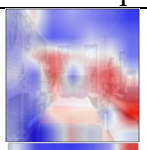
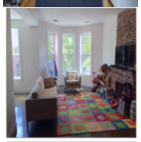
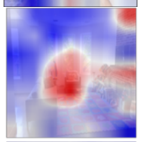
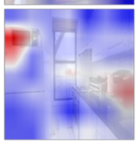
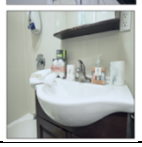
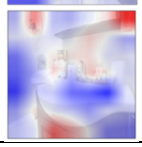


NOTE: The basic units of a neural network (e.g., Conv2D) are standard and not specific to our contrastive learning; therefore, they are not discussed. For details, please refer to He et al. (2016).

The heat map allows us to pinpoint how specific pixel-level features of an image contribute to its predicted uniqueness score. Table 1 provides examples for demonstration, including the model-predicted uniqueness score, the original image, and the RELAX-generated heat map. In these heat maps, red indicates a positive influence on the predicted uniqueness, while blue indicates a negative impact, with intensity representing the magnitude of the effect. For example, the bedroom image's high uniqueness score is attributed primarily to areas with features such as the bookshelves on either side of the bed, the orange blanket on the bed, and the bicycle on the wall. The living room's moderate uniqueness is concentrated in the area of the centered armchair and the mantel. The kitchen and bathroom images are deemed less unique with less red areas; whereas the blue color concentrates in areas of common appliances and toiletry, respectively.

It is necessary to note that our visual uniqueness measure is data-driven and operates from a bottom-up perspective, using the entire image to identify features that differentiate it from others without any predefined standards. This approach captures all differentiating features at the pixel level rather than predefined constructs (e.g., color contrast, objects). Thus, while our model reveals what makes each image unique, the contextual interpretation is left to the human viewer and can be subjective.

TABLE 1. EXAMPLES OF VISUAL UNIQUENESS HEAT MAP FROM RELAX

Room Type	Predicted Uniqueness	Original Image	Heat Map
Bedroom	14		
Living Room	10		
Kitchen	6		
Bathroom	2		

NOTE: Heat map colors represent different contribution levels to predicted uniqueness. Red indicates a positive contribution and blue a negative one, while darker shades signify greater magnitude and lighter shades less.

The interpretability of model predictions is a critical performance benchmark. By employing the RELAX framework, we have clearly identified which areas contribute to an image's uniqueness, translating complex model computations into visually intuitive heat maps. This level of interpretability not only validates the model's effectiveness but also provides actionable insights into managing visual uniqueness. Next, we report three experimental studies, both in the lab and online with human participants, to test the validity of our model in aligning with human perceptions of uniqueness.

MODEL VALIDATION

The objective of our machine learning model is to replicate human evaluation of visual uniqueness on a large scale. Ensuring that the model's predictions are in tune with human judgment is paramount. In line with established practices for validating measurement scales in marketing and consumer behavior research (Bearden, Netemeyer, and Haws 2011), we subject the model-derived uniqueness assessments to a battery of validity tests with human participants.

Study 1 is an eye-tracking experiment in which we examine the relationship between model-predicted uniqueness scores and human visual attention, testing the model's predictive or criterion validity. Study 2 is a lab experiment in which participants are asked explicitly to identify images they perceive as unique and to specify which features influenced their decisions, in order to examine the model's construct validity (i.e., whether it is measuring uniqueness as intended). Finally, Study 3 further tests construct validity by modifying images to remove features identified by the RELAX framework as contributing to uniqueness, observing if and how these changes affect human perceptions of uniqueness.

Study 1: Eye-Tracking Lab Experiment for Predictive Validity

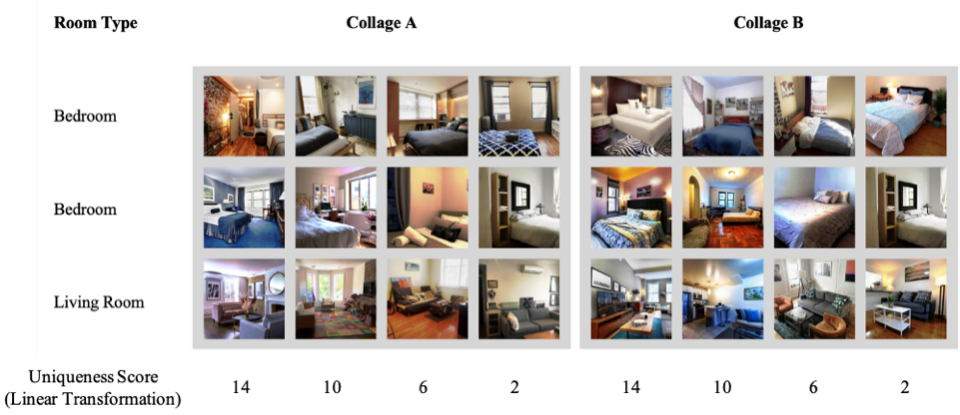
Previous literature suggests that stimuli that are unique tend to stand out and capture our attention naturally (Pieters, Warlop, and Wedel 2002; Theeuwes 2010). If our model sufficiently predicts visual uniqueness, there should be a positive correlation between the machine-predicted uniqueness scores and human visual attention. To test this, we recruited 212 university students in Europe for an eye-tracking study in a behavioral lab equipped with eye-tracking glasses.

Procedure. Upon arriving at the lab, the participants were fitted with eye-tracking glasses, which were then calibrated for accuracy. They then stood in front of a screen to view 12 Airbnb property images—eight bedroom and four living room images actually used by Airbnb

properties; this bedroom-to-living room ratio closely mirrors that of typical Airbnb properties. Each participant saw one of two image sets: Collage A or Collage B (see figure 4). We selected images representing four levels of model-predicted uniqueness scores—2, 6, 10, and 14. We randomized the positions of the images across four display variations to minimize the influence of image placement on attention—a factor that can affect consumer behavior (Sevilla and Townsend 2016). Additionally, we normalized the image brightness to avoid attention biases caused by unusual illumination.

The participants were asked to look through these Airbnb images for a trip to NYC. We tracked and analyzed participants' eye fixation—periods when the gaze is steady, indicating focused attention (Coppola and Purves 1996)—across the images as a measure of visual attention (Wedel and Pieters 2000). Two key metrics were recorded: total duration of fixations, indicating the length of time a participant's gaze lingered on a point; and fixation count, indicating the total number of fixations on an image (Pieters, Warlop, and Wedel 2002). These metrics reflect different aspects of information processing: while fixation count suggests the volume of information processed (Wedel and Pieters 2000), the total duration of fixation relates to the type and complexity of the processed information (Rayner 1998).

FIGURE 4. COLLAGES A AND B USED IN STUDIES 1 AND 2



Results. We first conducted a correlation analysis to test the alignment between model-predicted uniqueness scores and human attention at the image level. As presented in table 2, for both the total duration of fixations (in seconds; Mean = 155.30, SD = 626.01) and fixation count (Mean = .30, SD = 1.45), we observe positive and significant correlations with model-predicted visual uniqueness at the image level: corr. = .65 for the former ($p < .001$) and corr. = .60 for the latter ($p = .002$). The suggest that images deemed unique by our model garner more human attention, consistent with previous research predictions (Pieters, Warlop, and Wedel 2002).

TABLE 2. CORRELATION BETWEEN MODEL-PREDICTED VISUAL UNIQUENESS AND HUMAN ATTENTION

Human attention measure	**Correlation w/ uniqueness at image-condition-level**	***t***	***p***	**Correlation w/ uniqueness at image-level**	***t***	***p***
Total duration of fixations	.457	4.98	< .001	0.650	4.01	<.001
Fixation count	.364	3.79	< .001	0.603	3.55	.002

NOTE: There are 12 images in eight different conditions; thus, the degree of freedom in the correlation test at the image-condition level is 94. There are in total 12 images in each collage, and thus the degree of freedom in the correlation test at the image-level is 22.

Next, we employ a mixed-effect regression (Abebe et al. 2024) to isolate the impact of image uniqueness on human attention, detailed in Equation (1):

$$Human\ Attention_{ik} \sim \beta_0 + \beta_1 Model\ Uniqueness_k + \beta_2 Aesthetics_k + \beta_3 Complexity_k + Participant_i + P_k + \varepsilon_{ik}, \quad (1)$$

where *k* represents an image and *i* represents a participant. We use model-predicted uniqueness ($Model\ Uniqueness_k$) to predict the dependent variable, human attention ($Human\ Attention_{ik}$), for *i* observing k. The model incorporates random intercepts for participants ($Participant_i$), hence accommodating correlations in attention attribution within participants. We controlled for variables that might influence attention, such as image aesthetic quality ($Aesthetics_k$), visual

complexity ($Complexity_k$) computed by feature congestion (Ptak et al. 2009), and image position fixed effects (P_k). Finally, ε_{ik} is the error term.

Table 3 reports the results from Equation (1). Image aesthetic quality did not significantly predict attention ($\beta_2 = 164.15$, $p = .28$), implying that uniqueness is not merely a function of aesthetics. Visual complexity has a negative coefficient ($\beta_3 = -103.65$, $p = .03$), suggesting that visually simpler images tend to receive more attention, consistent with previous findings (Berlyne and Lawrence 1964). When controlling for image aesthetics and complexities, model-predicted uniqueness has a significant effect on both the total duration of fixations ($\beta_1 = 40.74$, $p < .001$) and fixation count ($\beta_1 = .07$, $p < .001$).[15] Lastly, it is essential to clarify that we do not conclude that participants focus on an image solely because of its uniqueness. Instead, participants naturally exhibited attention patterns consistent with the presence of unique images, proving the model's ability to identify elements of visual uniqueness. In Study 2, we formally demonstrate the alignment between our model's predictions and human judgments.

TABLE 3. EFFECT OF MODEL-PREDICTED VISUAL UNIQUENESS SCORES ON HUMAN ATTENTION MEASURES

VARIABLES	ESTIMATES	S.E.
DV: Total Duration of Fixations		
Model Visual Uniqueness	**40.7358*** **	**(5.9224)**
Aesthetics	164.1465	(151.4993)
Visual Complexity	−103.6481**	(45.4618)
DV: Fixation Count		
Model Visual Uniqueness	**0.0668*** **	**(0.0125)**
Aesthetics	0.4064	(0.3188)
Visual Complexity	−0.1871**	(0.0957)
Image Position FE	Y	(Y)
Observations	2,455	
Number of groups (participants)	209	

[15] The difference in the scale of coefficients for *Model Uniqueness* across these measures is attributed to their differing scales (*Total Duration of Fixations*: Mean = 155.30 *vs*. *Fixation Count*: Mean = 0.30), not to disparate levels of explanatory power.

NOTE: Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. The image position fixed effect is included for both DVs.

Study 2: Lab Experiment for Construct Validity

Study 2 seeks to test our uniqueness measure's construct validity. We investigate two aspects: first, whether individuals perceive images with higher model-predicted uniqueness scores as more unique compared to those with lower scores; and second, whether the pixel-level image characteristics identified by our model align with human perceptions.

Procedure. We recruited 288 participants (54.86% female, 44.79% male, 0.34% other, average age 20.22) from a European university. In a computer lab, each participant was shown one of the two image sets (Collage A *vs*. Collage B used in Study 1) with all image placements randomized to avoid positional bias.

Participants were instructed to view 12 Airbnb listing images displayed on a computer screen, imagining that they were planning a trip to NYC. Their task was to select three images that they deemed most unique. Subsequently, for each selected image, participants were asked to identify and click on five specific points within the room that contributed to their perception of the image's uniqueness. We also collected participant demographics, including age and gender.

Results. We conduct two analyses to assess how well the model-predicted uniqueness scores align with human perceptions of uniqueness. The first analysis focuses on the relationship between model-predicted uniqueness scores and the likelihood of human participants selecting certain images as unique. We employ a mixed effect logit regression (Abebe et al. 2024), as outlined in Equation (2):

$$Selection_{ik} \sim \beta_0 + \beta_1 Model\ Uniqueness_k + \beta_2 Aesthetics_k + \beta_3 Complexity_k + Participant_i + P_k + \varepsilon_{ik}, \quad (2)$$

where k represents an image and i represents a participant. The dependent variable, $Selection_{ik}$, equals 1 if i selected k as the most unique and 0 otherwise. The model incorporates random intercepts for participants ($Participant_i$), hence accommodating correlations in image selection within participants; since each participant identifies three out of 12 images as the most unique, this selection automatically leaves the remaining nine unselected. As in Study 1, we controlled for variables that might influence selections, such as image aesthetic quality ($Aesthetics_k$), visual complexity ($Complexity_k$) computed by feature congestion (Ptak et al. 2009), and image position fixed effects (P_k). Finally, ε_{ik} is the error term. β_1 measures the impact of the model-predicted uniqueness score on the likelihood that participants select an image as unique. The marginal effect of logistic regression in table 4 highlights that images with higher model-predicted uniqueness scores are also more likely to be selected as unique by human participants, as confirmed by the positive and significant coefficient of *Model Visual Uniqueness*. Specifically, a one-unit increase in model-predicted uniqueness (approximately a 6% increase in uniqueness) leads to a 4.12% increase in selection likelihood, even when controlling for various factors like image aesthetics and complexity. Since each image has a baseline probability of 30% of being selected,[16] this translates to a 14% increase in the probability of being selected as unique.

TABLE 4. MARGINAL EFFECT OF MODEL-PREDICTED VISUAL UNIQUENESS SCORES ON HUMAN SELECTION

VARIABLES	**DV: Selected as unique (binary variable)**	
	ESTIMATES	S.E.
Model Visual Uniqueness	**0.0412*** **	**(0.0031)**
Aesthetics	−0.4745***	(0.0911)
Visual Complexity	−0.0038	(0.0171)
Observations	3,455	
Number of groups (participants)	288	

[16] Thirty percent is computed as C(12, 2)/C(12, 3), where C(n, k) represents the number of combinations for selecting k instances from n cases without considering the order.

NOTE: Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$.

The second analysis compares the *uniqueness contribution heat maps* generated by the RELAX framework with the unique features identified by participants. Each image is divided into 64 × 64 sections or grids, and we analyze human clicks aggregated within each grid. Since grids are nested within each image that was previously selected as unique by a participant, we normalize the aggregated human clicks for each grid on the image by dividing them by the total clicks on the focal image. This normalization accounts for the likelihood of the image being selected in the first place. Equation (3) describes the model:

$$Relative\ Human\ Click\ \#_{kq} \sim \beta_0 + \beta_1 Model\ Importance_{kq} + \boldsymbol{\gamma} \cdot Image\ ID_k + \varepsilon_{kq}, \quad (3)$$

where k represents the image, and q (ranging from 1 to 64 × 64) denotes a specific grid on the focal image. The dependent variable, *Relative Human Click* $\#_{kq}$, denotes the total number of clicks by all participants on grid q of image k, normalized by the total number of clicks by all participants on image k. We control for image fixed effects, *Image* ID_k, to allow for image-specific factors to influence *Relative Human Click* $\#_{kq}$. The key variable, *Model Importance*$_{kq}$, is the sum of the model-predicted importance values at the pixel level for grid q of image k, normalized by the sum of importance values across all grids on k. Lastly, ε_{kq} is the random error.

The results, as presented in table 5, reveal a positive and significant correlation between the model-produced aggregated importance and the human-judged importance of image features. This suggests that our model accurately pinpoints areas contributing to an image's uniqueness in a way that aligns with human assessment.

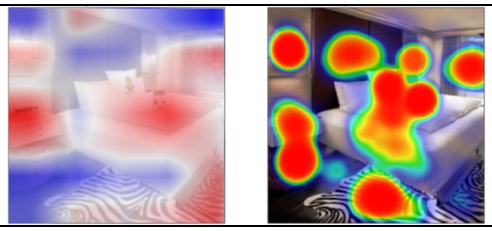
TABLE 5. REGRESSING HUMAN CLICKS ON MODEL-PREDICTED IMPORTANCE

Variables	Estimate	SD	*z*	*p*
Model-predicted Attribute Importance	**1.10**	**0.22**	**4.95**	**0.00**
Image ID-Fixed Effects		YES		
Observations		98,304		

NOTE: Dependent variable *Relative Human Click* $\#_{kq}$ denotes the total number of clicks of all participants on the *q*-th area of image *k*, divided by the total number of human clicks on image *k*. The coefficients of the image ID fixed effects are omitted due to space constraints; importantly, they are statistically insignificant. This suggests that the results are not affected by interdependencies among images, as we analyzed the normalized human clicks of each grid of each image.

In figure 5, we present a heat map comparison for a bedroom image, highlighting significant overlap between areas deemed important by human participants and those identified by our model. Notably, areas around key features, such as the plush toy on the bed, the curtain, and the lamp, exhibit a high degree of congruence.

FIGURE 5. AN EXAMPLE OF MODEL-PREDICTED HEAT MAP (LEFT) *VS.* AGGREGATED HUMAN CLICKS (RIGHT)



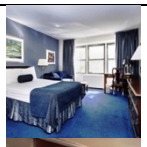
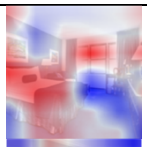
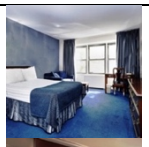
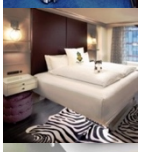
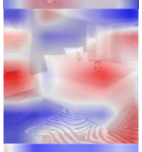
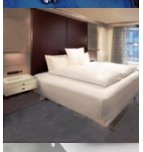
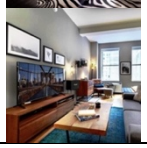
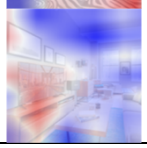
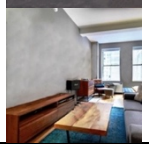
NOTE: The left panel heat map, predicted by the RELAX framework, uses color intensity to indicate the importance of each image area to its uniqueness score—red signifies a positive contribution, and blue indicates a negative contribution. The right panel heat map is generated based on our lab study data. It visualizes the total number of clicks each area received from participants; a deeper shade of red represents a higher concentration of clicks.

Study 3: Online Experiment with Manipulated Images

Building on the alignment between human assessments of uniqueness and the model predictions established in Study 2, Study 3 explores the practical implications of these findings. This study aims to experimentally modify images by removing features that our model identifies as key contributors to uniqueness predictions, in order to determine if these alterations impact human perceptions of image uniqueness.

Recall that the model-generated heat maps identify areas or pixels contributing positively to uniqueness with red shades, whereas darker intensities reflect greater magnitudes of contribution. The heat maps thus suggest how changes to image elements can affect uniqueness. For example, removing objects from prominently red areas is hypothesized to decrease the image's uniqueness score. As illustrated in table 6, removing objects highlighted in red, such as the pillow, lamp, and wall art, decreased the predicted visual uniqueness score of Image (a) from 14 to 4.4. Note in this study, we only removed elements highlighted in red to reduce uniqueness for simplicity. However, removing objects does not always reduce uniqueness; the predicted uniqueness score would increase if objects were removed from predominately blue areas (web appendix A).

TABLE 6. IMAGE MANIPULATION BASED ON MODEL-PREDICTED HEAT MAP

Image	Original Image	Original Uniqueness Score	Model-Predicted Heat Map	Manipulated Image	New Uniqueness Score/ Adjustments
(a)		14			4.4 *Removed pillow, lamps, wall arts*
(b)		14			2.5 *Removed toy, lamps, carpet, ottoman*
(c)		14			3.1 *Removed TV, books, wall art*

NOTE: New uniqueness scores are predicted based on the same model as the original images.

Procedure. This study recruited 399 participants across North America via Amazon Mechanical Turk (49.12% female, 50.88% male, average age 39.25). Participants viewed a set of six Airbnb property images arranged in a 3 × 2 collage. Participants were randomly assigned to one of four conditions: a control condition with no image manipulation, and three treatment conditions in which two of the three target images, shown in table 6, were altered to reduce their

predicted uniqueness by modifying specific image elements. The image collages used in each condition can be found in Web Appendix E.

Participants were asked to view a collage and imagine they were planning a trip to NYC, and then to choose the three images that they found the most unique out of the six. We collected participant demographic data, including age, gender, and pre-tax income, and prior travel experience, but did not use this information in our analysis.

Results. We assess whether modifying images suggested by our model-generated heat maps affects human judgments of uniqueness. Similar to Study 2, we employed a multi-level mixed effect logistic regression at the participant-image-condition level, involving 399 participants and three treated images each. As outlined in Equation (4), this analysis uses the manipulation condition to predict the likelihood of a participant selecting an image as unique:[17]

$$Selection_{ik} \sim \beta_0 + \beta_1 Manipulation_{k_i} + \beta_2 Visual\ Complexity_{k_i} + Participant_i + Condition_i + \varepsilon_{ik}, \quad (4)$$

where k (= 1, 2, 3) represents the image, and i (= 1, 2, …, 399) represents the participant. The dependent variable, $Selection_{ik}$, indicates whether image k was selected as unique by participant i. $Manipulation_{k_i}$ is a binary indicator of whether k was manipulated to decrease the uniqueness score for i. $Visual\ Complexity_{k_i}$ is the visual complexity (feature congestion; Ptak et al. 2009) of image k under the condition (which may or may not be manipulated; the visual complexity changes after manipulation) for i. The model includes the condition-level fixed effect $Condition_i$. Since participants made repeated choices, similar to Study 2, we control for participant random

[17] In a robustness test, we used the measured uniqueness score as the key independent variable instead of image manipulation and obtained similar results.

effects, $Participant_i$, to account for interdependencies among the selection of multiple images by the same participant. Lastly, ε_{ik} is the random error term.

Based on the marginal effect of logistic regression reported in table 7, image manipulation, which decreased the model-predicted uniqueness score, significantly reduced the likelihood of selection by 22.15% ($b = -0.2215$, $p < .001$). Similar to the calculations in Study 2, each image has a baseline probability of 75% of being selected, translating the effect of image manipulation to a 30% decrease (= 22.15% / 75%) in the probability of being selected as unique. This highlights that manipulating uniqueness-contributing elements, as identified by the uniqueness heat maps, changes the perceived visual uniqueness. Notably, the coefficient of visual complexity is statistically insignificant, suggesting that changes in this factor resulting from object removal do not significantly influence perceptions of visual uniqueness.

TABLE 7. MARGINAL EFFECT OF IMAGE MANIPULATION ON HUMAN SELECTION

VARIABLES	**DV: Selected as unique (binary variable)**	
	Estimates	S.E.
Manipulation	**−0.2215*** **	**(0.0375)**
Visual Complexity	−0.0390	(0.0331)
Condition FE	YES	
Observations	1,197	
Number of groups	399	

NOTE: Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$.

Discriminant Validity of Visual Uniqueness

We confirm the discriminant validity of visual uniqueness from concepts such as aesthetics, visual complexity, and atypicality.

First, we used the NIMA algorithm (Talebi and Milanfar 2018; introduced in Study 1) and extracted aesthetic scores for each Airbnb property image in our dataset. We then calculated the Pearson correlation between the uniqueness scores and the aesthetic scores across all images and

by room type. This analysis yields a negligible correlation between image uniqueness and aesthetics (corr. = −.01, $p < .001$), suggesting that these are separate attributes. Further analysis of cover images also revealed no significant correlation between uniqueness and aesthetics (corr. = −.0022, $p = .70$). This reinforces that visual uniqueness, as defined and measured by our unsupervised contrastive model, operates independently of aesthetic appeal.

Similarly, we test the discriminant validity of our model-predicted visual uniqueness from visual complexity by computing their correlations. We draw on prior studies and operationalized visual complexity using four measures: 1) the quantity of objects in the image, identified using a Mask R-CNN[18] (He et al. 2017), 2) the dissimilarity of objects, estimated via subband entropy (Machado et al. 2015), 3) the irregularity of object arrangement, approximated with feature congestion (Overgoor et al. 2022; Rosenholtz, Li, and Nakano 2007), and 4) subjective visual complexity, measured by edge density—the percentage of pixels that are edge pixels (Mack and Oliva 2004), calculated by Canny's edge detection algorithm (Canny 1986). We find no or very low correlation between our visual uniqueness scores and object count (corr. = 0.0088, $p = 0.16$), subband entropy (corr. < .0001, $p = .84$), feature congestion (corr. = .0002, $p = .90$), or edge density (corr. = .0142, $p < .001$), suggesting that our visual model-predicted uniqueness is distinct from visual complexity. This distinction is crucial for understanding what contributes to the perception of uniqueness in an image, beyond mere complexity or clutter.

Finally, there are both conceptual and empirical differences between visual uniqueness and atypicality. Although both concepts involve an image's distinctiveness from others, "atypicality"

[18] We use a subset due to the computational complexity of the Mask RNN model. Specifically, processing 30,000 images on an NVIDIA V100 GPU, known for its high-performance three-core parallel computing, took over 10 hours. This prolonged processing time can be attributed to several factors: First, the Mask RNN model's intricate architecture includes multiple stages such as object detection, classification, and instance segmentation, each requiring substantial computational resources. Second, the model demands intensive computations for each image, particularly during instance segmentation, which involves predicting masks for every detected object. Third, the high resolution and large size of our Airbnb property images increase the computational load, thus slowing down the inference times.

is the opposite of "prototypicality"—defined as the distance between a focal image and the average of its category—and is operationalized accordingly (e.g., Landwehr, Labroo, and Herrmann 2011; Liu et al. 2017); in contrast, our conceptualization and operationalization of uniqueness rely on pairwise comparisons between individual images without calculating the category average. This is because we conceptualize uniqueness perception as a bottom-up process without knowing what a typical image of a category should look like. Empirically, we calculate atypicality of property images in our data[19] and uncover a minor but statistically negative correlation with uniqueness (corr. = -0.0087). See Web Appendix A for a more detailed discussion, including images with high uniqueness but low atypicality.

MODEL APPLICATION: AIRBNB DEMAND DATA

With a validated visual uniqueness measure, we assess the impact of visual uniqueness on property demand using Airbnb data. As previously discussed, visual uniqueness can influence property demand not only positively, potentially driven by attention capturing and experiential value (Berger and Heath 2007; Tian, Bearden, and Hunter 2001), but also negatively, due to its association with uncertainty (Bolton and Drew 1991; Ghose 2009). This tension suggests a complex, nuanced relationship between property demand and visual uniqueness.

Past literature has documented an inverted U-shaped relationship between product demand and product features such as novelty, complexity, and prototypicality. Berlyne (1970) adapted

[19] We calculate the atypicality based on the Euclidean distance between the focal image and the centroid on all of our property image data using 512-dimensional vectors extracted from the ResNet model pretrained on ImageNet; note that this method still takes all image information into consideration, which is different from the traditional feature-based prototype calculation (Landwehr, Wentzel, and Herrmann 2013; Liu et al. 2017).

the Wundt curve to predict an inverted U-shaped relationship between stimuli novelty/complexity and hedonic liking, noting that complex stimuli became more pleasant as they became less novel. Meyers-Levy and Tybout (1989) found that a moderate deviation from a mental schema is optimal for product evaluations. Liu et al. (2017) identified an inverted U-shaped relationship between car design prototypicality and market share.

Drawing upon the extant literature and the tension discussed above, we predict an inverted U-shaped relationship between visual uniqueness and demand on Airbnb. Additionally, due to the potential increase in uncertainty perception associated with uniqueness, we explore how factors such as host responsiveness and property ratings might moderate the uniqueness–demand relationship. Drawing on Campbell and Goodstein's (2001) findings that (moderately) incongruent designs are preferred when perceived risk is low or absent, we predict that a high host response rate and positive property ratings can reduce booking risks, thereby moderating the impact of visual uniqueness on demand. This moderation could enhance the benefits of uniqueness, especially for highly unique properties.

Data Description

We constructed an *Airbnb Property Image Dataset* by collecting images monthly from Airbnb listings in NYC from April 2022 to April 2023. In parallel, we obtained historical data on demand, pricing, and property characteristics for these listings over the same period from AirDNA,[20] a third-party data provider, which is referred to as the *Airbnb Demand Dataset*. Following Zhang, Lee, Singh, et al. (2022), we removed stale properties, outliers, and missing data resulting from data scraping issues. Integrating the two datasets yields a single dataset of

[20] See website of AirDNA at https://www.airdna.co/. Previous research on Airbnb also use data from AirDNA (Zhang, Lee, Singh, & Srinivasan, 2022).

14,121 properties across 13 months in NYC. Details on the sample construction are presented in Web Appendix B. Table 8 provides definitions of the key variables and their summary statistics.

TABLE 8. DESCRIPTION AND SUMMARY STATISTICS OF KEY VARIABLES

Variables	Description	Mean	SD
Image-Level Visual Uniqueness	Visual uniqueness scores (linear transformation) for all 481,747 property images.	12.81	1.87
Property-Level Visual Uniqueness	Visual uniqueness scores averaged at the property-month level across property images listed for that month.	12.85	0.58
Location	Zip code areas and counties in NYC.		
Listing Type	Private room (39.44%), shared room (1.21%), entire home/apartment (59.35%).		
Property Type	In total, 40 types, including rental units (73.45%), homes (8.69%), townhouses (4.80%), condos (4.48%), lofts (2.65%), etc.		
Bedrooms	Number of bedrooms.	1.29	0.88
Bathrooms	Number of bathrooms.	1.14	0.43
Amenity Count	Number of amenities listed on the property detail page, which can be selected by the host before listing their properties and filtered by consumers in the search engine. Amenities that can be chosen from include pool, gym, elevator, TV, wireless, aircon, heating, free parking, breakfast, family friendly, washer, dryer, kitchen, smoking allowed/prohibited, free parking, suitable for events, etc.	31.45	12.14
Price (USD)	Average daily rate in a month.	206.44	169.43
Occupancy Rate	The portion of days that a property was booked in a month out of the days that the property was open for booking.* If there are no open days, then the observation is removed.	0.76	0.29
Number of Reviews	Accumulated number of reviews listed on the webpage from registration of the focal property until the focal month.	90.81	87.10
Overall Rating	Overall lodging experience rated by guests on a scale of 0–100.	94.70	6.14
Response Rate	The percentage of new inquiries and reservation requests that the host responded to (either accept, pre-approve, or decline) within 24 hours in the past 30 days.	95.65	13.22
Airbnb Superhost	A Superhost is a special badge that recognizes a host's superior performance. Every quarter, the Airbnb platform evaluates hosting performance over the past 12 months for all listings on a host's account.	0.20	0.40
Maximum Guests	Maximum number of guests allowed.	3.15	2.08
Minimum Stay	Minimum number of days for booking.	8.84	17.19
Instant Book Enabled	Whether a host allows guests to book immediately instead of needing to send a request for approval.	0.21	0.40
Number of Images	Number of property images listed for the focal month	18.16	11.10
Comparable Property Supply	"Comparable" units are those with the same number of bedrooms. The zip code-level supply is measured as the number of days listed by all comparable units in the same	1332.28	1488.53

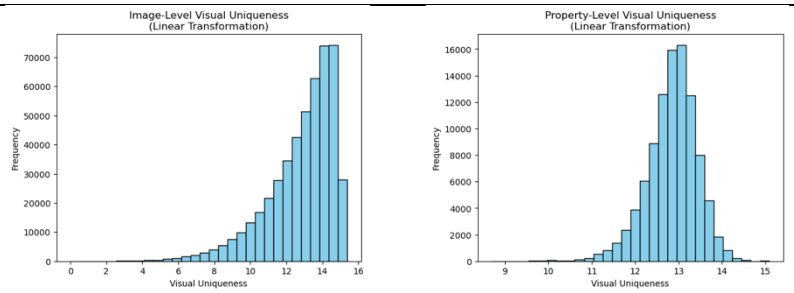
zip code area. This follows Li, Kim, and Srinivasan (2022).

NOTE: For categorical features, including Location, Listing Type, and Property Type, summary statistics are not provided. For dummy variables, such as Airbnb Superhost and Instant Book Enabled, we transform 0 and 1 into numerical values and calculate the corresponding mean and SD.

* There are times when a property is unavailable due to the host blocking it, such as for cleaning or when a friend is visiting.

Property Uniqueness Score. We apply our contrastive machine learning model to each image to derive a uniqueness score. For each property each month, we then calculate an average uniqueness score based on all the images listed for that property. Figure 6 illustrates the comparison between the distributions of image-level uniqueness scores vs. property-level uniqueness scores (averaged over time). Additionally, the geographical distribution of Airbnb property uniqueness scores, aggregated by zip code, is available in Web Appendix F.

FIGURE 6. IMAGE-LEVEL VS. PROPERTY-LEVEL VISUAL UNIQUENESS SCORES



(a) Image-Level Uniqueness (b) Property-Level Uniqueness

NOTE: The property-level uniqueness scores were averaged over time. As in Studies 1–3, the uniqueness score is computed with a linear transformation of contrastive loss L (see figure 2). The other two transformations, exponential and reciprocal, yield similar patterns. Details are in Web Appendices C and F.

Empirical Specification and Addressing Endogeneity

Our empirical approach closely follows that of Li, Kim, and Srinivasan (2022), who used a hedonic regression model on Airbnb longitudinal demand data. This model decomposes Airbnb listing prices into their constituent attributes, such as location, property type, amenities, and host characteristics. It also accounts for how pricing and occupancy rates are established, thus acknowledging property heterogeneity (Pauwels et al. 2022). The hedonic models of Airbnb price (P) and occupancy rate (D) for property j in month t are as follows:

$$P_{jt} = \alpha_0 + \alpha_1 X_{jt} + \alpha_2 H_j + \alpha_3 S_{jt} + \tau_{ct}^{p} + \varepsilon_{jt}^{p}, \text{ and} \quad (5)$$

$$D_{jt} = \beta_0 + \beta_1 U_{jt} + \beta_2 U_{jt}^2 + \beta_3 P_{jt} + \beta_4 X_{jt} + \beta_5 H_j + \beta_6 S_{jt} + \tau_{ct}^{d} + \varepsilon_{jt}^{d}, \quad (6)$$

where P_{jt} represents the log-transformed average daily rate of property j at time t; D_{jt} denotes the demand of property j at time t, calculated as the log of the occupancy rate. The independent variables include X_{jt} (property features), H_j (household demographics),[21] S_{jt} (the Airbnb supply of comparable units[22] of property j in the same zip-code area in month t), and U_{jt} (the visual uniqueness of property j at time t). Following Li, Kim, and Srinivasan (2022), we control for county-specific year-and-month fixed effects, τ_{ct}^{d}, to capture market-specific seasonal patterns in prices and demand across five counties (Bronx, Kings, New York, Queens, and Richmond) over a period of 13 months. We also control for visual aesthetics (NIMA model; see Talebi and Milanfar 2018) and visual complexity (feature congestion; Rosenholtz, Li, and Nakano 2007).

Addressing Endogenous Variables. Some variables may be endogenous, for example, price and image uniqueness; they are endogenous if hosts set higher prices and choose more unique images simultaneously to reflect the property's perceived value. We deal with the endogenous price using the average property characteristics (i.e., bathroom count) of noncomparable Airbnb properties in the same zip code as instruments. Noncomparable properties in the same zip code are those with different numbers of bedrooms (Li, Kim, and Srinivasan 2022).[23] In addition, to

[21] Following the same practice by Li, Kim, and Srinivasan (2022), we collected from zip-code-level demographics from the American Housing Survey and impute the host demographics for the Airbnb properties using the local zip-code-level demographics, including age, income, percent with children, percent married, and education. Data available at U.S. Census Bureau 2021 (https://www.census.gov/programs-surveys/acs/).

[22] "Comparable" units are those with the same number of bedrooms. The zip code-level supply is measured by the number of days listed by all comparable units in the same zip code area (Li, Kim, and Srinivasan 2022).

[23] These are predetermined and unlikely to be correlated with unobserved demand shocks. For example, price might be an endogenous variable as there exist common shocks, such as pandemics or interest rate cuts, that correlate with both price and occupancy rate, but we cannot control for them due to a lack of data over time. In such instances, the use of the mean attributes of noncomparable Airbnb properties may effectively mitigate the endogeneity concern associated with pricing, as the characteristics of competitors are presumed to exhibit minimal correlation with the latent demand shocks affecting a property. Nevertheless, the competitive landscape, particularly the proximity of a product's characteristics to those of its rivals, exerts a substantial impact on market competition, subsequently influencing product markup and pricing dynamics (Berry, Levinsohn, and Pakes 1995; Berry 1994).

deal with potential endogeneity arising from unobserved seasonal shocks correlated with demand or supply, we employ BLP-type instruments (Berry, Levinsohn, and Pakes 1995; Berry 1994). Following precedent in research using Airbnb data (Li, Kim, and Srinivasan 2022; Li and Srinivasan 2019), we deal with the endogeneity of supply by using the Airbnb supply of noncomparable properties as the instrumental variable. Web Appendix F provides detailed discussion on potential endogeneity sources related to image uniqueness, including price, local competition, location, and market seasonal factors, along with our approaches to addressing them. Lastly, the price and occupancy rate (Equations [5] and [6]) are estimated simultaneously using a three-stage least squares (3SLS) approach to allow for the correlation of the error terms in the two equations (Sargan 1964), in order to jointly model the interdependence between Airbnb pricing and occupancy.

Airbnb Demand Model Estimation Results

Table 9 presents the estimation results of two models: Column 1, which includes only the linear term of uniqueness score, and Column 2, which incorporates both linear and quadratic terms. The models account for various property features and supply variables, such as the number of property images, guest reviews, and comparable units in the same zip code (refer to table 8). Column 1 indicates that uniqueness has no significant impact on property demand on average. However, Column 2 reveals an inverted U-shaped relationship between uniqueness and demand, when controlling for the visual aesthetics and visual complexity of the property images. This finding supports our hypothesis that while an optimal level of uniqueness can enhance demand, excessively unique properties may deter potential customers, aligning with prior research on atypicality and incongruity. Figure 7 displays a marginal plot that illustrates the impact of image uniqueness on property demand.

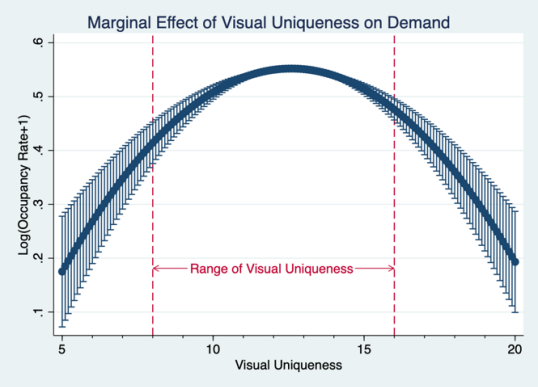
TABLE 9. LINEAR AND QUADRATIC EFFECTS OF UNIQUENESS ON DEMAND

	(1)	(2)
	D.V.: Property Demand	
VARIABLES	Linear Term of Uniqueness	Including Quadratic Terms of Uniqueness
Visual Uniqueness	**−0.00103**	**0.165*** **
	(0.000914)	**(0.0225)**
(Visual Uniqueness)^2		**−0.00654*** **
		(0.000889)
Property Daily Rate	−0.173***	−0.172***
	(0.0109)	(0.0109)
Response Rate	0.0538***	0.0541***
	(0.00402)	(0.00402)
Overall Rating	0.333***	0.332***
	(0.0162)	(0.0162)
Visual Aesthetics	0.00471	0.00479
	(0.00652)	(0.00651)
Visual Complexity	0.00374***	0.00377***
	(0.00128)	(0.00128)
Number of Images	0.0148***	0.0130***
	(0.00135)	(0.00137)
Number of Reviews	0.0214***	0.0215***
	(0.000991)	(0.000991)
Airbnb Superhost	0.00448***	0.00444***
	(0.00135)	(0.00135)
Number of Max Guests	0.112***	0.111***
	(0.00922)	(0.00921)
Instant Book Enabled	−0.0163***	−0.0161***
	(0.00134)	(0.00133)
Minimum Stay	2.36e-06	5.19e-07
	(4.14e-05)	(4.14e-05)
Number of Room Amenities	0.0339***	0.0332***
	(0.00222)	(0.00222)
Bedrooms	0.0105***	0.0106***
	(0.00122)	(0.00122)
Bathrooms	0.0179***	0.0178***
	(0.00145)	(0.00145)
Median Age	0.000483**	0.000474**
	(0.000198)	(0.000198)
Median Income	−1.22e-06***	−1.22e-06***
	(3.95e-08)	(3.95e-08)
Children	0.127***	0.123***
	(0.0345)	(0.0345)
Married	−0.0344*	−0.0308
	(0.0207)	(0.0207)

Bachelors	0.428***	0.426***
	(0.0119)	(0.0119)
Supply Count	0.00285***	0.00301***
	(0.000806)	(0.000806)
Fixed Effects	County-Month	County-Month
Observations	95,673	95,673
R-squared	0.242	0.243

NOTE: The D.V. is *log(occupancy rate+1)*. Robust standard errors are reported in parentheses. *Supply Count* refers to the number of comparable units in the same zip code, defined as those with the same number of bedrooms, and is log-transformed. The zip code-level supply is calculated based on the total days listed by all comparable units in that area (Li, Kim, and Srinivasan 2022). Significance levels: * < .1; ** < .05; *** < .01.

FIGURE 7. MARGINAL EFFECT OF VISUAL UNIQUENESS ON DEMAND



NOTE: The plot expands from 5 to 20 to show the overall trend; however, the range of average visual uniqueness in our sample at the property level is from approximately 8 to 16, as illustrated between the red dashed lines.

How substantial is the impact of visual uniqueness according to the results? The coefficients of the linear and quadratic terms of *Visual Uniqueness* translate to a change of $0.1650 \times Unique - 0.0065 \times (Unique)^2$ on monthly *log(occupancy rate+1)*. Therefore, when uniqueness positively influences demand (as seen on the rising slope of the curve in figure 7), a one-standard-deviation increase in the property-level uniqueness score results in a 15.9% increase in property demand. This translates to an additional yearly revenue of $9,809.70, assuming that the property remains open throughout the year and that the increase is sustained over that period.[24]

[24] Revenue is approximated as $(e^{.1585} - 1) \times Occupancy\ Rate \times \frac{Days}{Year} \times Daily\ Rate\ (USD) = (e^{.1585} - 1) \times .7580 \times 365 \times 206.44\ (USD) \approx 9809.74\ (USD)$. Note that this was calculated based on a one-standard-deviation

One might wonder if the observed increase in property demand is due to Airbnb's recommendation algorithm favoring properties with higher visual uniqueness. Without access to the proprietary details of Airbnb's algorithm, we cannot conclusively determine this. However, to assess the potential impact on our findings, we analyzed search data from Airbnb. We conducted searches across various price ranges and check-in/check-out dates, analyzing 270 search results spread over 15 pages for each query. We then examined the relationship between the visual uniqueness of a property and its position in the search results. Our analysis revealed no significant correlation between these two factors, that is, Airbnb's algorithm does not appear to prioritize visual uniqueness in its recommendations. Further details of this analysis are available in Web Appendix F.

Moderation Effects on Visual Uniqueness

Prior literature suggests that product characteristics that confer uniqueness to offerings can introduce layers of uncertainty or risks (Ghose 2009), potentially tempering the positive impact of uniqueness on demand. In a P2P market like Airbnb, where the listed offerings are not standardized, the uncertainty is particularly high (Mao and Lyu 2017). Thus, we explore the moderating role of uncertainty mitigation; we conjecture that when uncertainty factors are alleviated—such as through a high host response rate or property rating—the potential negative effects associated with a high degree of uniqueness may be lessened.

A host's response reflects the host's reliability and service quality, while the overall property rating captures the accommodation quality as judged by past guests. Consequently, high

(1 SD = 0.58) increase in uniqueness. The 0.11 in within-property uniqueness variance was believed to be unintentional. However, although this natural variance is small, it represents an additional $1,675 in yearly revenue. Although smaller, this incremental economic benefit is not negligible.

property ratings and a high host response rate might reduce uncertainty and moderate the possible drawbacks of visual uniqueness. We test these two variables as proxies for accommodation quality and uncertainty to see how they moderate the effects of visual uniqueness on property demand. Both variables, scaled from 0% to 100%, are divided by 100 for analysis. Equation (7) details the model, including interactions between these moderators and visual uniqueness:

$$D_{jt} = \beta_0 + \beta_1 U_{jt} + \beta_2 U_{jt}^2 + \gamma U_{jt} \times M_j + \beta_3 P_{jt} + \beta_4 X_{jt} + \beta_5 H_j + \beta_6 S_{jt} + \tau_{ct}^m + \varepsilon_{jt}^d, \quad (7)$$

where the elements remain the same as in Equation (6), except for M_j, which denotes the moderator. We used a 3SLS method for joint estimation.

As shown in table 10, Columns 1 and 2 examine the moderating role of host response rate, while Columns 3 and 4 explore the moderating effect of property overall rating. As can be seen, both variables positively and significantly moderate the linear and quadratic terms of visual uniqueness. Additionally, as depicted in figure 8, an increased response rate or overall rating tends to flatten the U-shaped uniqueness-demand relationship. These findings suggest that high quality of accommodation, indicated by a high response rate or overall rating, which reduces uncertainty, can enhance the positive impact of uniqueness on attracting guests.

We illustrate the economic impact of moderation results by comparing two hypothetical properties with identical characteristics (both at the sample mean) but differing in their average uniqueness scores: Property A, with a uniqueness score of 10, is on the upward slope of the inverted U-shaped curve, while Property B, with a score of 14, is on the downward slope. Given that the mean of the inverse U-shape is around 12, the effect of uniqueness on demand for Property A is higher than that for Property B. If both increase their host response rate (scaled) from 80% to 90%, the projected percentage changes in demand would be 0.0016 for Property A and 0.0069 for Property B, translating to annual revenue increases of $91.40 vs. $394.10,

respectively. Similarly, improving their overall property rating (scaled) change from 0.8 to 0.9 could boost their annual revenue by $291.30 for Property A and $2,658.80 for Property B. This also suggests that the moderation effect of both moderators flatten the inverse U-shaped curve.

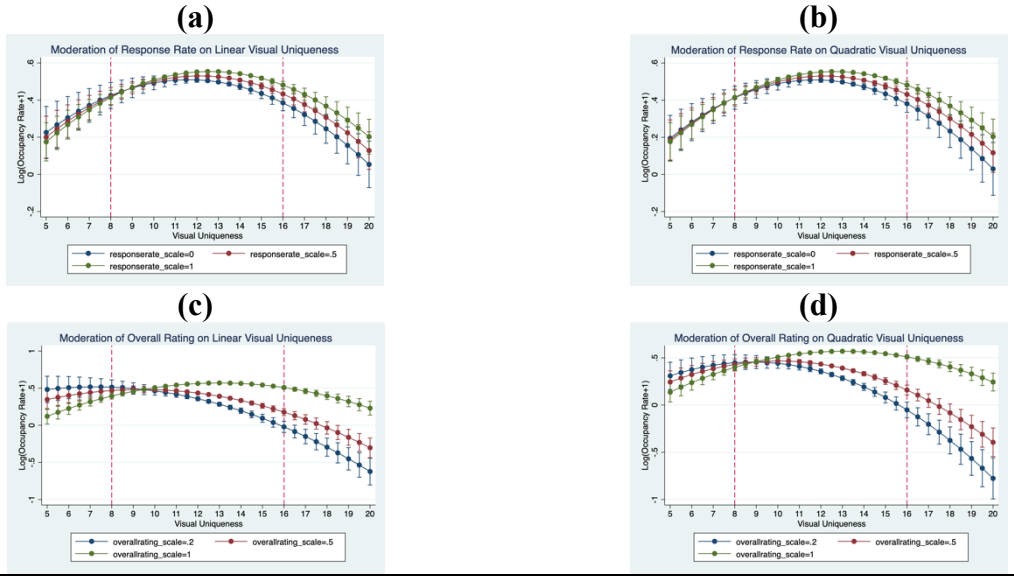
TABLE 10. MODERATION EFFECTS OF HOST AND PROPERTY INFERENCE

	(1)	(2)	(3)	(4)
	D.V.: Property Demand			
VARIABLES	Moderator: Response Rate (Scaled)		Moderator: Overall Rating (Scaled)	
Visual Uniqueness	0.151***	0.164***	0.0805***	0.175***
	(0.0235)	(0.0225)	(0.0252)	(0.0226)
(Visual Uniqueness)^2	−0.00649***	−0.00698***	−0.00698***	−0.0107***
	(0.000889)	(0.000915)	(0.000890)	(0.00105)
Visual Uniqueness × Moderator	**0.0133****		**0.101*****	
	(0.00635)		**(0.0136)**	
(Visual Uniqueness)^2 × Moderator		**0.000514****		**0.00398*****
		(0.000250)		**(0.000538)**
Property Daily Rate	−0.172***	−0.172***	−0.171***	−0.171***
	(0.0109)	(0.0109)	(0.0109)	(0.0109)
Response Rate	−0.117	−0.0310	0.0534***	0.0534***
	(0.0817)	(0.0416)	(0.00402)	(0.00402)
Overall Rating	0.331***	0.331***	−0.959***	−0.315***
	(0.0162)	(0.0162)	(0.172)	(0.0873)
Visual Aesthetics	0.00492	0.00491	0.00532	0.00531
	(0.00651)	(0.00651)	(0.00651)	(0.00651)
Visual Complexity	0.00378***	0.00378***	0.00383***	0.00383***
	(0.00129)	(0.00129)	(0.00128)	(0.00128)
Number of Images	0.0130***	0.0130***	0.0127***	0.0127***
	(0.00137)	(0.00137)	(0.00137)	(0.00137)
Number of Reviews	0.0215***	0.0215***	0.0216***	0.0216***
	(0.000993)	(0.000993)	(0.000987)	(0.000987)
Airbnb Superhost	0.00443***	0.00443***	0.00411***	0.00412***
	(0.00135)	(0.00135)	(0.00135)	(0.00135)
Number of Max Guests	0.111***	0.111***	0.111***	0.111***
	(0.00924)	(0.00924)	(0.00919)	(0.00920)
Instant Book Enabled	−0.0161***	−0.0161***	−0.0161***	−0.0161***
	(0.00133)	(0.00133)	(0.00133)	(0.00133)
Minimum Stay	8.29e-07	8.06e-07	−2.49e-06	−2.51e-06
	(4.14e-05)	(4.14e-05)	(4.14e-05)	(4.14e-05)
Number of Room Amenities	0.0330***	0.0330***	0.0328***	0.0328***
	(0.00223)	(0.00223)	(0.00221)	(0.00221)
Bedrooms	0.0107***	0.0107***	0.0107***	0.0107***

	(0.00122)	(0.00122)	(0.00122)	(0.00122)
Bathrooms	0.0178***	0.0178***	0.0178***	0.0178***
	(0.00145)	(0.00145)	(0.00145)	(0.00145)
Median Age	0.000483**	0.000483**	0.000481**	0.000481**
	(0.000198)	(0.000198)	(0.000197)	(0.000197)
Median Income	−1.22e-06***	−1.22e-06***	−1.22e-06***	−1.22e-06***
	(3.95e-08)	(3.95e-08)	(3.94e-08)	(3.94e-08)
Children	0.125***	0.125***	0.126***	0.126***
	(0.0345)	(0.0345)	(0.0345)	(0.0345)
Married	−0.0313	−0.0313	−0.0310	−0.0311
	(0.0207)	(0.0207)	(0.0207)	(0.0207)
Bachelors	0.426***	0.426***	0.425***	0.425***
	(0.0119)	(0.0119)	(0.0119)	(0.0119)
Supply Count	0.00305***	0.00305***	0.00311***	0.00310***
	(0.000805)	(0.000805)	(0.000806)	(0.000806)
Fixed Effects	County-Month	County-Month	County-Month	County-Month
Observations	95,673	95,673	95,673	95,673
R-squared	0.243	0.243	0.244	0.244

NOTE. The D.V. is *log(occupancy rate+1)*. Robust standard errors are reported in parentheses. *Supply Count* refers to the number of comparable units in the same zip code, defined as those with the same number of bedrooms, and is log-transformed. The zip code-level supply is calculated based on the total days listed by all comparable units in that area (Li, Kim, and Srinivasan 2022). Significance levels: * < .1; ** < .05; *** < .01.

FIGURE 8. MARGINAL EFFECT OF VISUAL UNIQUENESS ON DEMAND AT DIFFERENT LEVELS OF MODERATORS (RESPONSE RATE, OVERALL RATING)



NOTE: The x-axis represents visual uniqueness, and the y-axis measures Airbnb property demand. The four figures illustrate the linear and quadratic interactions between visual uniqueness and a moderator. Two moderators, host response rate and overall property rating, are scaled by 1/100 for clearer interpretation. The curves are color-coded based on the moderator level: 0 (blue), 0.5 (red), or 1 (green). Figure (a) plots the marginal moderation effect of the host response rate on the linear term of visual uniqueness. Figure (b) depicts the marginal moderation effect of

the host response rate on the quadratic term of visual uniqueness. Figure (c) displays the marginal moderation effect of the overall property rating on the linear term of visual uniqueness. Figure (d) shows the marginal moderation effect of the overall property rating on the quadratic term of visual uniqueness. Each plot expands from 5 to 20, yet the range of average visual uniqueness at the property level, in our sample is approximately 8 to 16, as illustrated between the red dashed lines.

Additional Analyses

In Web Appendix F, we detail a series of additional analyses testing the robustness of our main results and exploring other ways uniqueness may affect demand.

First, we use the number of reservation days as an alternative demand measure and applied alternative transformations (exponential and reciprocal) to convert contrastive loss into uniqueness scores. Similar results are obtained. Next, we separately control for the atypicality of the images and the linear and quadratic terms of visual complexity measured by edge density in the demand regression. We find that the effect of uniqueness on demand is still significant in an inversed U-shape when controlling for atypicality and visual complexity, validating the independence of visual uniqueness from these concepts. Additionally, we analyze how distributional aspects of visual uniqueness—the median, variance, minimum, and maximum uniqueness scores of images within a property—affect demand. Our results suggest that properties with images displaying lower variance and less extremes in uniqueness scores enjoy higher consumer demand. This preference may stem from reduced uncertainty in quality perception, which aligns with findings on the benefits of consistency in advertising (Aaker 2012; Keller et al. 2010). Finally, we investigate the differential impacts of cover versus non-cover images. As shown in table W12 of the web appendix, when including both uniqueness scores, only the average uniqueness of non-cover images shows a significant impact: displaying an inverted-U-shaped effect on demand. The lack of significant influence by cover image uniqueness alone may be attributed to the deliberateness of the accommodation selection

process: Instead of making immediate choices based on the unique cover images, it is more likely that consumers evaluate the complete set of property images and other property characteristics to make the final accommodation choice.

GENERAL DISCUSSION

Our research embarks on a comprehensive journey to understand and quantify the concept of visual uniqueness within the burgeoning landscape of P2P marketplaces, particularly Airbnb. The study uses contrastive loss and random data augmentation techniques to develop an unsupervised machine learning model trained and tested on a large-scale image dataset of Airbnb properties. We enhance the current understanding of what makes an image unique by interpreting the contrastive learning model post hoc and generating uniqueness heat maps using the RELAX framework, which pinpoints key features at the pixel level that contribute to a property's uniqueness. These insights were rigorously validated through three distinct controlled experiments, including an eye-tracking study, to ensure alignment with human perception. Applying the validated unsupervised machine learning model to a longitudinal demand dataset of Airbnb properties, we uncovered an inverted U-shaped relationship between visual uniqueness and property demand. The findings further suggest that properties with higher host response rates or superior quality gain more substantial benefits from visual uniqueness.

Theoretical and Substantive Contribution

This research significantly advances the current theoretical understanding of visual uniqueness, especially in P2P marketplaces like Airbnb. It merges insights from marketing, cognitive psychology, and machine learning to quantify a qualitative and subjective concept.

First, our work broadens the traditional conceptualization of product differentiation that is rooted in feature-based category prototypes (e.g., Landwehr et al. 2011). Observing that

idiosyncratic differences in P2P offering images constitute their uniqueness, we propose a uniqueness perception model that takes all image information into account. Grounded in cognitive psychology theories (Berlyne 1970; Gibson and Gibson 1955; Itti and Koch 2001; Regehr and Brooks 1993; Theeuwes 2010), we conceptualize the perception and evaluation of visual uniqueness in environments with diverse offerings as a bottom-up, comparative, and holistic process in which all image features are compared between images without relying on preconceived schemas or standardized features. This is crucial in P2P settings, where traditional prototype models fall short of capturing nuanced differentiation. This aligns with the notion that seemingly irrelevant attributes are important in creating product differentiation (Carpenter, Glazer, and Nakamoto 1994). Our model can accurately predict human judgments of visual uniqueness, validating our conceptualization and deepening the current understanding of visual uniqueness in highly varied and nuanced stimuli.

Second, our work joins and expands the dialogue on the effect of visual characteristics on consumer choice. On the one hand, we establish visual uniqueness as an independent construct distinguishable from related constructs (e.g., aesthetic appeal, atypicality, and visual complexity), underscoring its explanatory power above and beyond these factors. On the other hand, we identify an inverted U-shaped relationship between visual uniqueness and property demand and further demonstrate that properties with high ratings and high host responsiveness benefit more from visual uniqueness. While previous research has predicted similar relationship patterns, such as the effect of prototype deviation (Berlyne 1974; Hekkert, Snelders, and Van Wieringen 2003; Liu et al. 2017) and visual complexity on preference (Berlyne 1970; Berlyne and Lawrence 1964), our work complements this literature by providing large-scale empirical evidence on the effect of visual characteristics on actual product demand.

Lastly, we add to the emerging literature on peer-to-peer marketplaces by demonstrating the unique contribution of visual uniqueness on demand. P2P marketplaces are characterized by their idiosyncrasy of offerings and those who offer them (Li and Lutz 2019). Recent research has examined how some dimensions of this idiosyncrasy, such as the quality of Airbnb property images (Zhang et al., 2021) and whether hosts smile in their profile photos (Zhang et al. 2025), affect demand. Our findings underscore the significant role of visual uniqueness, a quality that captures idiosyncrasies that visually differentiate one offering from others, in influencing consumer attention, perception, and demand within P2P marketplaces.

Methodological Contribution

The present research demonstrates that unsupervised machine learning models constructed without the need for labeled human-input samples can successfully approximate human perceptions. This integration of knowledge from consumer research with advancements in artificial intelligence (AI) has led to three significant methodological contributions.

First, we are among the first to use psychological theories in constructing and validating unsupervised machine learning models. Supervised machine learning models have been used in marketing research, where a large number of features are measured and added to a predictive model to predict key dependent variables (Yarkoni and Westfall 2017). Psychological insights may be used for predictor selection or model fine-tuning following the supervised approach; however, such an approach is ineffective when the dependent variables are unobservable or too costly to obtain. Instead, we leverage psychological insights to develop a machine learning model that simulates, to some extent, human cognitive processes. This enables us to effectively reduce dependency on human-labeled data while still predicting human judgments accurately.

This verifies the critical role played by deep consumer knowledge in guiding the development of AI tools to predict and influence consumer behaviors and market responses at increasing speeds.

Second, the use of eye-tracking experiments and validation studies involving human participants to systematically validate a machine learning model is an innovative approach. Our study goes beyond demonstrating that unsupervised machine learning models *can* align with human judgments; it also elucidates *how* such alignment can be tested. We adapted validation techniques traditionally used in psychology for scale development (Bearden, Netemeyer, and Haws 2011; Campbell and Fiske 1959) to assess the construct validity of our machine learning model. This provides a new roadmap for future research aiming to capture complex psychological constructs using AI technologies.

Furthermore, this research brings innovation to the emerging applications of machine learning in consumer research. Machine learning is increasingly being harnessed by marketing researchers to yield consumer insights based on extensive, unstructured image data (Wang, Bendle, and Pan 2024). Examples include studies on how image quality affects consumers' accommodation choices (Zhang et al. 2022), how visual components of logos influence brand identity (Dew, Ansari, and Toubia 2021), and what facial features improve one's chance of becoming a celebrity (Feng et al. 2025). The development of an unsupervised machine learning model that utilizes contrastive learning (LeCun and Huang 2005) represents a significant advancement in the field. This model not only quantifies visual uniqueness, but does so in a way that reflects human judgment, thereby overcoming the limitations of existing models that rely heavily on labeled data. We provide an appropriate case for the application of contrastive learning and representation learning techniques (Becker and Hinton 1992) in image analytics in

marketing research. The innovative application of explainable AI techniques, such as the creation of uniqueness heat maps, further strengthens the model's practical utility and interpretability.

Finally, our study demonstrates the feasibility and efficiency of conducting "synthetic consumer research" using unsupervised machine learning. Traditionally, consumer research relies on experiments involving human participants to predict responses. However, our model can simulate these predictions on a broader scale. This approach aligns with the emerging interest in using large language models to expand consumer behavior research (Brand, Israeli, and Ngwe 2024; Horton 2023). Through empirical validation, we show that our model-generated uniqueness scores and uniqueness heat maps for Airbnb property images closely match human judgments. This enables us to assess visual uniqueness, without needing to conduct costly surveys or experiments with large samples of human participants.

Managerial Implications

Our research offers critical insights for operators and providers in P2P marketplaces. The ability to measure and understand visual uniqueness at scale provides a valuable tool for enhancing the appeal of listings on platforms such as Airbnb. It allows for the strategic presentation of properties to maximize their attractiveness to potential customers, balancing uniqueness with broader appeal to optimize market performance.

Leveraging uniqueness in marketing and recommendations. P2P platforms can enhance their marketing strategies by highlighting the unique aspects of properties. Showcasing properties that exemplify the right level of uniqueness can help attract consumer attention. This can be particularly effective in digital marketing campaigns and social media, where visual impact is crucial. For platforms like Airbnb, incorporating a visual uniqueness metric into search and recommendation algorithms could improve the user experience. The platform could also

selectively highlight properties with higher uniqueness scores for users seeking novel experiences, thereby enhancing customer satisfaction.

Balancing uniqueness with quality guarantees. Our discovery of an inverted U-shaped relationship between uniqueness and demand suggests that hosts should avoid overemphasizing uniqueness. Hosts should strike a balance, ensuring that their properties are unique enough to stand out, but not so idiosyncratic that they alienate potential guests. Specifically, properties that are further away from the optimum value of visual uniqueness based on our demand regression (i.e., 12.61) can benefit more in leveraging our model to increase demand, as the slope of the inverse U-shape between uniqueness and demand is much steeper away from the optimum. Also, hosts offering highly unique offerings should thrive to provide quick responses and quality guarantees to mitigate perceived uncertainties for prospective guests to reap the maximum benefit of uniqueness. Moreover, keeping image uniqueness consistent—avoiding extremely high or low scores—within the property image set also help hosts appeal to consumers.

Actionability for hosts. Imagine if Airbnb implemented an image analysis tool using our model, new hosts, those renovating properties, or those updating images could use insights from this tool to guide investments in renovations, furnishings, and decorations. Using uniqueness scores and heat maps, hosts can achieve desirable property visual uniqueness by either adjusting the image set or altering physical room features. Specifically, removing the least unique image from a set increases the average uniqueness across listings, as our data shows a 0.53 SD increase in average property-level uniqueness after dropping the least unique image from each property. Additionally, hosts can optimize visual uniqueness by modifying room elements without major spatial changes. Using a uniqueness heat map, hosts can identify which room features contribute positively or negatively to visual uniqueness. For example, changing visual elements in negative

areas marked by dark blue can enhance uniqueness, whereas disturbing features in positive areas marked by dark red may decrease it. However, since uniqueness and its contributing image features are determined at the individual image level, hosts must employ a trial-and-error approach to identify the effective strategies for optimizing their property's visual uniqueness.

Application to other contexts. Our methodological framework has broad applicability across various domains where images are abundant and idiosyncratic, and the visual uniqueness of images may pose significant impacts on consumer behaviors. This includes, but is not limited to, social media platforms such as Instagram, P2P real estate marketplaces like Zillow, and e-commerce platforms like Amazon. It is important to clarify, however, that while our framework is versatile, this does not imply that our current model predictions can be directly used in these contexts without modifications. To adapt our model to a different context, it would be necessary to obtain an image dataset from that context and train a new model following the same methodological steps, with the same model structure but fine-tuned parameters.

Limitations and Future Research

This research is not without limitations. *First*, while robust, our machine learning model is optimized on property images of Airbnb listings in NYC. These contextual specificities limit us from making broader generalizations about how uniqueness affects product demand in other contexts. Future studies could apply the model to different platforms or types of offerings to examine the relationship between market performance and visual uniqueness. *Second*, this work focuses on the visual aspect of uniqueness, which may not provide a full picture. Combining image data with textual reviews or user profiles could offer a more holistic view of how visual and descriptive elements interact to influence consumer preferences. For example, future research could examine whether the images and elements driving up visual uniqueness scores are

mentioned in reviews as making the product more unique, thereby improving customer satisfaction. *Third*, our research examines how visual uniqueness impacts property bookings at the property level, yet a more granular, image-level analysis—exploring how visual uniqueness influence guest engagement such as clicks, views, or live chats—could provide deeper insights. While our focus on property-level analysis aligns with our study objectives, it necessitates data aggregation, which inevitably sacrifices some finer details. This broader approach, nonetheless, effectively captures the primary variables and controls relevant to our study. Although detailed interaction data is unavailable, supplementary analyses suggest that individual image uniqueness, including that of the cover image, does not significantly affect demand in isolation. Our current data limitations preclude a more granular examination, but future research could productively investigate how the uniqueness of a cover image alone influences click-through rates—a metric highly relevant in contexts like advertising, where first impressions are crucial. It would also be fruitful to study how different visual uniqueness scores of the same product, the distribution of these scores, and the order of the presentation influence consumer viewing experience and their subsequent choices. *Lastly*, although we delineated a plausible conceptual model for visual uniqueness perception and validated the computational model built in accordance with it, direct testing of the new theory is beyond our scope. Future research can test whether uniqueness perception in visually complex environments is indeed bottom-up, comparative, and holistic.

In conclusion, this research marks a significant stride in understanding and quantifying visual uniqueness within P2P marketplaces. By establishing robust framework grounded in psychological theories and leveraging unsupervised machine learning models without requiring labeled samples, our research demonstrates the feasibility of approximating human perceptions computationally. Additionally, we illustrate that machine learning models can be systematically

and rigorously validated for their construct, predictive, and discriminant validity, establishing their practical and theoretical robustness. This work underscores the potential for integrating complex human psychological constructs with AI technologies, offering a systematic approach for developing and validating computational models that bridge the gap between human cognition and machine learning.

DATA COLLECTION STATEMENT

The data is scraped from Airbnb website and purchased from AirDNA company. The dats is collected from April 2022 - April 2023. Shunyuan Zhang and Xiaohang (Flora) Feng collected the data. Shunyuan Zhang, Charis Li, and Xiaohang (Flora) Feng analyzed the data. The data is currently stored on HBSGrid.

REFERENCES

Gossa Fetene Abebe, Menen Tilahun, Hana Tadesse, Abdu Seid, Tariku Yigremachew, Anteneh Messele Birhanu, Desalegn Girma (2024), "Predictors of Delayed Initiation of Breast Milk and Exclusive Breastfeeding in Ethiopia: A Multi-Level Mixed-Effect Analysis," *PLoS One*, 19 (4), e0301042.

William O. Bearden, Richard G. Netemeyer, and Kelly L. Haws (2011), *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research*, Thousand Oaks, CA, US: SAGE Publications, Inc.

Becker, Suzanna and Geoffrey E. Hinton (1992), "Self-Organizing Neural Network that Discovers Surfaces in Random-Dot Stereograms," *Nature*, 355 (6356), 161–63.

Berger, Jonah and Chip Heath (2007), "Where Consumers Diverge from Others: Identity Signaling and Product Domains," *Journal of Consumer Research*, 34 (2), 121–34.

Daniel Ellis Berlyne (1970), "Novelty, Complexity, and Hedonic Value," *Perception & Psychophysics*, 8 (5), 279–86.

Daniel Ellis Berlyne (Ed.) (1974), Studies in the New Experimental Aesthetics: Steps Toward an Objective Psychology of Aesthetic Appreciation, Oxford, England: Hemisphere.

Daniel Ellis Berlyne and George H. Lawrence (1964), "Effects of Complexity and Incongruity Variables on GSR, Investigatory Behavior, and Verbally Expressed Preference," *The Journal of General Psychology*, 71 (1), 21–45.

Berry, Steven T. (1994), "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25 (2), 242–62.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4), 841.

Bloch, Peter H. (1995), "Seeking the Ideal Form: Product Design and Consumer Response," *Journal of Marketing*, 59 (3), 16–29.

Bolton, Ruth N. and James H. Drew (1991), "A Multistage Model of Customers' Assessments of Service Quality and Value," *Journal of Consumer Research*, 17 (4), 375.

Bouvier, Seth E. (2009), "Top-Down Influences of Spatial Attention in Visual Cortex," *The Journal of Neuroscience*, 29 (6), 1597–98.

James Brand, Ayelet Israeli, Donald Ngwe (2024), "Using GPT for Market Research," working paper.

Campbell, Donald T. and Donald W. Fiske (1959), "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 56 (2), 81–105.

Campbell, Margaret C. and Ronald C. Goodstein (2001), "The Moderating Effect of Perceived Risk on Consumers' Evaluations of Product Incongruity: Preference for the Norm: Table 1," *Journal of Consumer Research*, 28 (3), 439–49.

Canny, John (1986), "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8 (6), 679–98.

Carpenter, Gregory S., Rashi Glazer, and Kent Nakamoto (1994), "Meaningful Brands from Meaningless Differentiation: The Dependence on Irrelevant Attributes," *Journal of Marketing Research*, 31 (3), 339.

Cesareo Ludovica, Townsend Claudia, Pavlov Eugene (2023), "Hideous but Worth It: Distinctive Ugliness as a Signal of Luxury," *Journal of the Academy of Marketing Science*, 51 (3), 636–57.

Sumit Chopra, Raia Hadsell, and Yann LeCun (2005a), "Learning a Similarity Metric Discriminatively, with Application to Face Verification," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 539–46.

Chung, Jaeyeon (Jae), Gita Venkataramani Johar, Yanyan Li, Oded Netzer, and Matthew Pearson (2022), "Mining Consumer Minds: Downstream Consequences of Host Motivations for Home-Sharing Platforms," *Journal of Consumer Research*, 48 (5), 817–38.

Coppola, David and Dale Purves (1996), "The Extraordinarily Rapid Disappearance of Entopic Images," *Proceedings of the National Academy of Sciences*, 93 (15), 8001–4.

Dellaert, Benedict G. C. (2019), "The Consumer Production Journey: Marketing to Consumers as Co-Producers in the Sharing Economy," *Journal of the Academy of Marketing Science*, 47 (2), 238–54.

Dew, Ryan, Asim Ansari, and Olivier Toubia (2021), "Letting Logos Speak: Leveraging Multiview Representation Learning for Data-Driven Branding and Logo Design," *Marketing Science*.

Feng, Xiaohang, Zhang, Shunyuan, Liu, Xiao, Srinivasan, Kannan, & Lamberton, Cait. (2025). An AI Method to Score Celebrity Visual Potential. Journal of Marketing Research, (forthcoming). https://doi.org/10.1177/00222437251323238.

Ghose (2009), "Internet Exchanges for Used Goods: An Empirical Analysis of Trade Patterns and Adverse Selection," *MIS Quarterly*, 33 (2), 263.

Gibson, James J. (2002), "A Theory of Direct Visual Perception," in *Vision and Mind: Selected Readings in the Philosophy of Perception*, Cambridge, MA, US: Boston Review, 77–91.

Gibson, James J. and Eleanor J. Gibson (1955), "Perceptual Learning: Differentiation or Enrichment?," *Psychological Review*, 62 (1), 32–41.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2016), "Deep Residual Learning for Image Recognition," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 770–78.

He, Kaiming, Georgia Gkioxari, Piotr Dollar, and Ross Girshick (2017), "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, 2980–88.

Hekkert, Paul, Dirk Snelders, and Piet C. W. Van Wieringen (2003), "'Most Advanced, Yet Acceptable': Typicality and Novelty as Joint Predictors of Aesthetic Preference in Industrial Design," *British Journal of Psychology*, 94 (1), 111–24.

Horton, John J. (2023), "Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?," NBER Working Paper, Cambridge, MA.

Industry Research Biz (2023), "Global Sharing Economy Industry Research Report 2023, Competitive Landscape, Market Size, Regional Status and Prospect."

Itti, Laurent and Christof Koch (2001), "Computational Modelling of Visual Attention," *Nature Reviews Neuroscience*, 2 (3), 194–203.

Kang, Sungmin and Youn Kue Na (2020), "Effects of Strategy Characteristics for Sustainable Competitive Advantage in Sharing Economy Businesses on Creating Shared Value and Performance," *Sustainability*, 12 (4), 1397.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton (2017), "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, 60 (6), 84–90.

Landwehr, Jan R., Aparna A. Labroo, and Andreas Herrmann (2011), "Gut Liking for the Ordinary: Incorporating Design Fluency Improves Automobile Sales Forecasts," *Marketing Science*, 30 (3), 416–29.

Lauga, Dominique Olié, Elie Ofek, and Zsolt Katona (2022), "EXPRESS: When and How Should Firms Differentiate? Quality and Advertising Decisions in a Duopoly," *Journal of Marketing Research*, 002224372210820.

Le-Khac, Phuc H., Graham Healy, and Alan F. Smeaton (2020), "Contrastive Representation Learning: A Framework and Review," *IEEE Access*, 8, 193907–34.

LeCun, Yann and Fu Jie Huang (2005), "Loss Functions for Discriminative Training of Energy-Based Models," *Proceedings of Machine Learning Research*, 206–13.

Li, Charis X. and Richard J. Lutz (2019), "Object History Value in the Sharing Economy," in *Handbook of the Sharing Economy*, Edward Elgar Publishing.

Li, Hui and Kannan Srinivasan (2019), "Competitive Dynamics in the Sharing Economy: An Analysis in the Context of Airbnb and Hotels," *Marketing Science*, 38 (3), 365–91.

Li, Hui, Yijin Kim, and Kannan Srinivasan (2022), "Market Shifts in the Sharing Economy: The Impact of Airbnb on Housing Rentals," *Management Science*, 68 (11), 8015–44.

Lipton, Zachary C. (2018), "The Mythos of Model Interpretability," *Queue*, 16 (3), 31–57.

Liu, Yan, Krista J. Li, Haipeng Chen, and Subramanian Balachander (2017), "The Effects of Products' Aesthetic Design on Demand and Marketing-Mix Effectiveness: The Role of Segment Prototypicality and Brand Consistency," *Journal of Marketing*, 81 (1), 83–102.

Penousal Machado, Juan Romero, Marcos Nadal, Antonino Santos, João Correia, Adrián Carballal (2015), "Computerized Measures of Visual Complexity," *Acta Psychologica*, 160, 43–57.

Shie Mannor, Xin Jin, Jiawei Han, Xinhua Zhang (2011), "K-Means Clustering," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 563–64.

Mao, Zhenxing and Jiaying Lyu (2017), "Why Travelers Use Airbnb Again?," *International Journal of Contemporary Hospitality Management*, 29 (9), 2464–82.

Mack, Mack L. Oliva, Aude. (2004). Computational estimation of visual complexity. Annual Object, Perception, Attention, and Memory Conference. Minneapolos, Minnesota.

McMains, Stephanie and Sabine Kastner (2011), "Interactions of Top-Down and Bottom-Up Mechanisms in Human Visual Cortex," *The Journal of Neuroscience*, 31 (2), 587–97.

Meyers-Levy, Joan and Alice M. Tybout (1989), "Schema Congruity as a Basis for Product Evaluation," *Journal of Consumer Research*, 16 (1), 39.

Oord, Aaron van den, Yazhe Li, and Oriol Vinyals (2018), "Representation Learning with Contrastive Predictive Coding."

Overgoor, Gijs, William Rand, Willemijn van Dolen, and Masoud Mazloom (2022), "Simplicity is not key: Understanding firm-generated social media images and consumer liking." *International Journal of Research in Marketing*, 39 (3), 639-655.

Palmeri, Thomas J. and Isabel Gauthier (2004), "Visual Object Understanding," *Nature Reviews Neuroscience*, 5 (4), 291–303.

Pauwels, Koen, Bharat Sud, Robert Fisher, and Kersi Antia (2022), "Should You Change Your Ad Messaging or Execution? It Depends on Brand Age," *Applied Marketing Analytics: The Peer-Reviewed Journal*, 8 (1), 43–54.

Pieters, Rik, Luk Warlop, and Michel Wedel (2002), "Breaking Through the Clutter: Benefits of Advertisement Originality and Familiarity for Brand Attention and Memory," *Management Science*, 48 (6), 765–81.

Pieters, Rik, Michel Wedel, and Rajeev Batra (2010), "The Stopping Power of Advertising: Measures and Effects of Visual Complexity," *Journal of Marketing*, 74 (5), 48–60.

Ptak, Radek, Laetitia Golay, René M. Müri, and Armin Schnider (2009), "Looking Left with Left Neglect: The Role of Spatial Attention When Active Vision Selects Local Image Features for Fixation," *Cortex*, 45 (10), 1156–66.

Rayner, Keith (1998), "Eye Movements in Reading and Information Processing: 20 Years of Research," *Psychological Bulletin*, 124 (3), 372–422.

Regehr, Glenn and Brooks, Lee R. (1993), "Perceptual Manifestations of an Analytic Structure: The Priority of Holistic Individuation," *Journal of Experimental Psychology*, 122 (1), 92–114.

Ruth Rosenholtz, Yuanzhen Li, Lisa Nakano (2007), "Measuring Visual Clutter," *Journal of Vision*, 7 (2), 17.

Sargan, John Denis (1964), “Three-Stage Least-Squares and Full Maximum Likelihood Estimates,” *Econometrica*, 32, 77–81.

Schroff, Florian, Dmitry Kalenichenko, and James Philbin (2015), “FaceNet: A Unified Embedding for Face Recognition and Clustering,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 815–23.

Pierre Sermanet, Corey Lynch, Yevgen Chebotar, Jasmine Hsu, Eric Jang, Stefan Schaal, Sergey Levine (2017), “Time-Contrastive Networks: Self-Supervised Learning from Video.”

Sevilla, Julio and Claudia Townsend (2016), “The Space-to-Product Ratio Effect: How Interstitial Space Influences Product Aesthetic Appeal, Store Perceptions, and Product Preference,” *Journal of Marketing Research*, 53 (5), 665–81.

Shorten, Connor and Taghi M. Khoshgoftaar (2019), “A Survey on Image Data Augmentation for Deep Learning,” *Journal of Big Data*, 6 (1), 60.

Snyder, Charles R. and Howard L. Fromkin (1977), “Abnormality as a Positive Characteristic: The Development and Validation of a Scale Measuring Need for Uniqueness,” *Journal of Abnormal Psychology*, 86 (5), 518–27.

Talebi, Hossein and Peyman Milanfar (2018), “NIMA: Neural Image Assessment,” *IEEE Transactions on Image Processing*, 27, 3998–4011.

Theeuwes, Jan (2010), “Top–Down and Bottom–Up Control of Visual Selection,” *Acta Psychologica*, 135 (2), 77–99.

Kelly Tepper Tian, William O. Bearden and Gary L. Hunter (2001), “Consumers’ Need for Uniqueness: Scale Development and Validation,” *Journal of Consumer Research*, 28 (1), 50–66.

Wang, Xin, Neil Bendle, and Yinjie Pan (2024), "Beyond Text: Marketing Strategy in a World Turned Upside Down," *Journal of the Academy of Marketing Science*.

Wedel, Michel and Rik Pieters (2000), "Eye Fixations on Advertisements and Memory for Brands: A Model and Findings," *Marketing Science*, 19 (4), 297–312.

Kristoffer K. Wickstrøm, Daniel J. Trosten, Sigurd Løkse, Ahcène Boubekki, Karl Øyvind Mikalsen, Michael C. Kampffmeyer, Robert Jenssen (2021), "RELAX: Representation Learning Explainability."

Wu, Zhirong, Yuanjun Xiong, Stella Yu, and Dahua Lin (2018), "Unsupervised Feature Learning via Non-Parametric Instance-level Discrimination."

Yarkoni, Tal and Jacob Westfall (2017), "Choosing Prediction Over Explanation in Psychology: Lessons from Machine Learning," *Perspectives on Psychological Science*, 12 (6), 1100–22.

Yin, Nan, Li Shen, Mengzhu Wang, Long Lan, Zeyu Ma, Chong Chen, Xian-Sheng Hua, and Xiao Luo (2023), "CoCo: A Coupled Contrastive Framework for Unsupervised Domain Adaptive Graph Classification."

Zhang, Shunyuan, Dokyun Lee, Param Singh, and Tridas Mukhopadhyay (2022), "Demand Interactions in Sharing Economies: Evidence from a Natural Experiment Involving Airbnb and Uber/Lyft," *Journal of Marketing Research*, 59 (2), 374–91.

Zhang, Shunyuan, Dokyun Lee, Param Singh, and Kannan Srinivasan (2022), "What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features," *Management Science*, 68 (8), 5644–66.

Shunyuan Zhang, Elizabeth M S Friedman, Kannan Srinivasan, Ravi Dhar, Xupin Zhang, Serving with a Smile on Airbnb: Analyzing the Economic Returns and Behavioral Underpinnings of the

Host's Smile, *Journal of Consumer Research*, Volume 51, Issue 6, April 2025, Pages 1073–1097, https://doi.org/10.1093/jcr/ucae049.

Zhong, Yuanyi, Haoran Tang, Junkun Chen, Jian Peng, and Yu-Xiong Wang (2022), "Is Self-Supervised Learning More Robust Than Supervised Learning?"

CHAPTER IV: SUSTAINABILITY AND COMPETITION ON AMAZON

ABSTRACT

We examine the impact of Amazon‘s Climate Pledge Friendly (CPF) badge on market dynamics. We applied a game-theoretic model and causal inference using data from Amazon.com to explore the effects of this badge on consumer behavior, seller pricing, and market concentration. Our theoretical model outlines a three-stage process in which sellers set prices, the marketplace determines badge eligibility, and consumers make purchase decisions. We discovered that increased demand, higher prices, and reduced market concentration occur when the green consumers with positive attitudes toward the badge outnumber non-green consumers with a negative attitudes toward it. Our model showed that certifying only the most sustainable products led to the highest prices, strongest demand, and lowest market concentration, outperforming badge strategies that certified all or no products. Empirically, we gathered six months of data on 6,606 products across eight categories, using the interactive fixed effect counterfactual (IFEct) estimator to manage endogeneity and treatment reversals. Our findings indicate that the CPF badge significantly enhances sales volume, increases product prices, and decreases market concentration. These results guide sellers considering green certification and platforms contemplating unified green badge policies.

Keywords: sustainability, eco-labeling, e-commerce, seller competition, causal inference, multimodal vector representation

INTRODUCTION

Sustainability and green marketing are significant trends, with sustainable product sales expected to grow by 19.5% from 2023 to 2032 (Global Market Insight). According to Forbes, in 2023, 90% of business leaders viewed sustainability as crucial, and 60% of companies had developed sustainability strategies. As eco-consciousness rises among consumers and businesses, marketplaces are incentivized to offer sustainability certifications. In 2020, reflecting its commitment to sustainability, Amazon introduced the Climate Pledge Friendly (CPF) badge, and as of September 2023, it had collaborated with 50 external certifiers and established two of its own certifications.[25] Any product sold with one of these 52 certifications qualifies for the CPF badge; thus, it is a unifying green badge that represents all 52 certifications. We define this policy as a unified green badge policy.

The impact of a green badge is complex. On one hand, rising eco-consciousness among consumers may boost sales, especially for brands that can compete more effectively in sustainability. On the other hand, a strong focus on sustainability might detract from product design, reducing its appeal and sales potential (White, Hardisty, and Habib 2019; Townsend and Shu 2010). As a result, the badge could hurt sales if consumers see the products as overpriced, of lower quality, or poorly designed. That is, consumers may be reluctant to pay a premium for sustainable products, and sustainable materials might compromise functionality, which can be viewed as a drawback (Luchs et al. 2010; Newman, Gorlin, and Dhar 2014). It follows that the badge's influence on pricing is also twofold: it could lead to higher prices if consumers are willing to pay more for green products, but it might lower prices if there is a gap between the stated willingness to pay and actual purchasing behavior or signify a more efficient way of

[25] See a detailed introduction at https://www.amazon.com/b?node=21221608011.

packaging or shipping the products, so firms could pass on part of the cost savings to consumers by lowering prices (Talwar et al. 2021). The effect of a green badge on market concentration remains underexplored, with limited research showing that tailoring products for different consumer segments could reduce competitors' profits (Amaldoss and Prusty 2024).

Given the complex effects of green badging on sales and pricing and the limited research on how they affect market concentration, this paper seeks to address this gap in the literature. Unlike previous research that has focused mainly on consumer-side outcomes, such as purchase intent or click-throughs, our paper adopts a supply-side perspective by analyzing seller pricing, market concentration, and small seller competitiveness. Proserpio et al. (2024), for example, identified the causal impact of the certified product feature on purchase behavior but devoted less attention to changes in market structure; Wang and He (2024) focused on how the CPF's information provision drives demand without exploring deeply how sellers might react. Our approach expands on prior work by examining how sellers respond to the CPF badge—both their own and those of competing sellers—and how these responses affect pricing, market concentration, and small-seller competitiveness. This supply-side perspective is critical because badges create strategic interactions among sellers, ultimately shaping market outcomes.

To capture the strategic interactions among sellers and consumers. we first developed a three-stage theoretical game theory model and documented a specific condition that facilitates a scenario in which the adoption of green badges leads to increased price and demand, coupled with reduced market concentration. This balance is achieved as long as the green segment of consumers with a positive attitude toward the badge is larger than the non-green segment of consumers with a negative attitude toward the badge. Our findings further suggest that an optimal balance occurs when the badging proportion is neither 0% or 100%, but instead falls

somewhere in between. This optimal scenario arises when the baseline utility of non-sustainability-related features for the badged product is at least twice that of the unbadged product. Additionally, the degree of product differentiation between the two sellers must be sufficiently large. Finally, the proportion of eco-conscious consumers with a positive perception of the badge should outweigh the proportion of non-green consumers who hold a negative perception.

Next, we offer empirical evidence of the multifaceted impacts of adopting a green badge in online marketplaces. We collected daily data spanning March 1, 2023 to September 15, 2023, resulting in 6,606 unique products across eight product categories. In our data, 35.39% of the products were badged, approximating the intermediate badge threshold, as predicted by our theoretical model. Thus, we hypothesized that the adoption of the CPF badge was associated with increased demand and price, as well as reduced market concentration.

To causally examine the impact of the badge on market outcomes, we used the interactive fixed effect counterfactual (IFEct) estimator (Liu, Wang, and Xu 2022). This approach is designed to handle treatment reversals (i.e., the treatment of a group is subsequently reversed or altered), and to address endogeneity concerns associated with unobserved counterfactuals through an augmented factor structure (Bai 2009). Treatment reversal exists in our dataset; 29.79% of the products received a CPF badge and then lost it at a later period (some might subsequently regain it). Further, to control for potential confounding variables, we included various time-varying covariates extracted from product reviews, descriptions, and images. We also employed machine learning models to derive insights from unstructured data such as product images and descriptions, enhancing the robustness of our estimates.

Our study presents three key findings. First, the CPF badge significantly increased overall product demand. Second, the badge was associated with a significant price increase, which is consistent with the literature on sustainable goods carrying a price premium (Tully and Winer 2014). Finally, the CPF badge enhanced the competitiveness of smaller brands, leading to reduced market concentration within subcategories, and fostering a more competitive environment for small brands on the platform.

By integrating insights from the game theory model with empirical evidence from Amazon sales data, we offer the mechanisms and necessary conditions that lead to increased demand, higher prices, and reduced market concentration following the adoption of a green badge. The theoretical model explains (1) why a badging policy (e.g., different levels of badge coverage) can lead to higher prices after CPF's introduction, (2) the market conditions under which these price effects occur, and (3) how changes in parameters (e.g., the proportion of green products) can alter seller behavior. By explicitly defining the players, their actions, and the information structure, our model illuminates mechanisms that might otherwise remain hidden in purely empirical work. Our empirical analysis then tests these theoretical predictions by tracking how CPF influences pricing patterns, market concentration, and the competitiveness of smaller sellers. The results confirmed that strategic seller responses aligned with our model's proposed mechanisms, underscoring the importance of a supply-side lens in understanding CPF's broader impact. Moreover, the model offers managerial guidance for platform operators—illustrating why badging works, how it shapes market structure, and when it may lead to higher prices.

This paper makes three substantive contributions: (i) On the *demand side,* we provide empirical evidence that CPF badge adoption leads to increased sales; (ii) on the *supply side*, we provide empirical evidence that CPF badge adoption leads to increased price and decreased

market concentration; and (iii) methodologically, we are among the first to integrate empirical causal inference with game theory modeling (Clark, Horstmann, and Houde 2024). We developed a game theory model explaining the sufficient conditions for increased demand, price, and market competition after implementing the unified green badge policy. However, we would like to highlight that the empirical evidence is our most important contribution, especially the findings on the supply-side effects on price and market competition, and the theoretical model was developed to explain why these results arise.

Our research represents a pioneering effort to examine the profound effects of a unified eco-label on demand, price, and market concentration across an e-commerce platform. We elucidate the mechanisms through which the adoption of such a unified green badge augments demand, elevates price levels, and reduces market concentration. This study not only clarifies the direct benefits of eco-labeling but also highlights its role in fostering a more competitive and sustainable marketplace.

LITERATURE REVIEW

We contribute to three strands of literature: the drivers of sustainability, the motivations behind brand sustainability marketing, and the impact of eco-labeling on product demand and pricing.

Drivers of Sustainable Consumption

Previous research on the drivers of sustainable consumption has primarily focused on product attributes. Research in this field often examines the nuanced impact of green attributes on product evaluations. Luchs et al. (2010) found that green products are often associated with attributes such as gentleness rather than strength, leading consumers to prefer conventional products when strength is a priority. Chen and Chang (2013) also noted that consumers

frequently perceived green products as less effective than their traditional counterparts. Additionally, Newman, Gorlin, and Dhar (2014) showed that when companies prominently highlight green attributes, consumers may assume that resources are being diverted from improving functionality, reducing perceptions of quality and purchase intentions. These studies highlight how the framing of environmental efforts can significantly influence consumer perceptions and preferences.

Motivations for Sustainability Marketing

The existing literature highlights three principal motivations for businesses to engage in sustainability marketing. First, companies that adapt to the evolving landscape, particularly by addressing the need for sustainability, position themselves for long-term survival and prosperity and gain strategic advantages (Banerjee, Iyer, and Kashyap 2003; Bollinger et al. 2024; Brecko and Kim 2024). Second, studies have consistently shown that adopting socially and environmentally responsible practices improves consumer perceptions and leads to greater profitability (Olsen, Slotegraaf, and Chandukala 2014; Sen and Bhattacharya 2001). Third, firms that embrace sustainable operations and pioneer innovative business models for sustainable consumption often achieve higher long-term profits, as demonstrated by successes in the sharing economy. While traditional marketing efforts have focused on identifying the green consumer, modern research emphasizes understanding the predictors of sustainable consumption (Menon and Menon 1997; Zhu 2024). This shift prompts marketers to expand their strategies, offering mutual long-term benefits to both businesses and the environment. As companies pursue more environmentally sustainable practices, it is necessary to cultivate recognition and rewards from consumers for the companies' sustainable values, potentially fostering sustainable consumption and enhancing the companies' sustainability and strategic positioning. Our contribution lies in

examining firms' potential profitability through pricing and content strategies under a unified green badge policy—an area largely underexplored due to the recency of such certification programs by platforms like Amazon.

The Impact of Eco-Labeling

Previous research has focused on how eco-labeling affects demand and pricing. Eco-labels communicate a product's sustainable attributes, helping consumers make informed eco-friendly choices (Parguel, Benoît-Moreau, and Larceneux 2011). However, information overload, limited exposure, and confusion can hinder sustainable behavior (Chen and Chang 2013). Influential labels that are attention-grabbing, easy to understand, and consistent across categories can better guide eco-friendly decisions. Contrasting positive labels with negative ones that highlight harmful attributes may further improve label effectiveness (Borin, Cerf, and Krishnan 2011).

The pricing of eco-labeling depends on consumers' willingness to pay a premium for sustainable products (Tully and Winer 2014). However, a gap often exists between stated willingness to pay and actual purchase behavior (Johnstone and Tan 2015). Third-party certification can enhance the transparency and credibility of eco-labels, but the proliferation of various green certificates may undermine their effectiveness due to concerns about credibility and certifying bodies (Borin, Cerf, and Krishnan 2011). Our study is among the first to explore the effects of a unified green badge, a consolidation of various green certificates, on sales and market concentration in e-commerce.

THEORETICAL MODEL

There were three motivations for constructing the theoretical analysis. First, because our paper's key contribution is the supply side analysis, the theoretical model can provide managerial guidance for the platform on why badging works and how it affects the market structure among

its sellers. A fundamental purpose of a theoretical model is to illuminate the causal mechanisms that shape individual and group behavior. By explicitly defining the players, their action sets, and the information structure, a game-theoretic framework clarifies the strategic motives that might otherwise remain obscure in purely empirical work. Rather than treating observed correlations as black-box relationships, a theoretical model explicitly articulates why certain strategic behaviors arise and how changes in parameters or constraints can alter outcomes.

Second, the theory model can identify necessary market conditions, including the platform's badging strategy, that lead to higher prices after implementation of the CPF badge. Without a theoretical foundation, it can be difficult to interpret why empirical results shift when environmental or institutional conditions change. Third, the theory model can generate testable hypotheses that can be tested with the data. In our theoretical model, we assume that both the focal and rival products hold external green certificates, allowing us to isolate the platform's badging decision from underlying product sustainability levels. In the empirical analysis, by contrast, we cover all listings (including non-green items) and therefore use only a binary CPF badge indicator without classifying products as green.

Sellers and Marketplace

Two competing sellers, A and B, offer substitutable green products in an online marketplace. Following previous research (Zhou and Zou 2023), we assume zero marginal production costs. However, the two sellers face different costs for developing sustainability features and obtaining external green certifications, which is consistent with Amazon's CPF program guidelines.[26] The two sellers determine their product prices, p_A and p_B, and pay the marketplace a percentage commission r ($0 < r < 1$) on the sale of their products. Hence, the

[26] An introduction to the cost associated with each external green certification can be found at https://m.media-amazon.com/images/G/01/rainier/help/Climate_Pledge_Friendly_Certification_Guide.pdf

marketplace's profits are proportional to the aggregated revenue generated by A and B. For a unit sale of product $j \in \{A, B\}$, the marketplace's profit is $r \times p_j$ and seller j's profit is $(1 - r) \times p_j$. The main analysis focuses on the case in which the marketplace's commission rate r is exogenous, which is consistent with the current industry practice of commission rates seldom changing, even if online marketplaces have experienced many changes.[27]

The marketplace decides whether to award the unified green badge to products based on its own sustainability standards. Suppose that the sustainability level $f_j \in (0, 1)$ of product j ($j \in \{A, B\}$), which is assessed according to the product's sustainability features (e.g., material) or factors in the product life cycle (e.g., carbon emissions during production), is exogenously given and will not change in a short time. Without loss of generalizability, we assume that product A has a high sustainability level, with $0.5 < f_A < 1$, while product B has a low sustainability level, with $0 < f_B < 0.5$. Given the varying levels of sustainability, the costs associated with developing sustainability features and obtaining external green certificates also differ. This variance is not solely because higher costs enable higher sustainability levels but also because each green certificate has distinct standards; consequently, the costs and membership fees for each certification vary.

Additionally, the marketplace selects a threshold, denoted by $I_0 \in [0,1]$, to selectively assign a badge to a product. The badging threshold is interpreted as the percentage of products receiving the badge. We further assume that as the badge threshold tightens, the cost of verifying a product's sustainability features rises—stricter standards demand more intensive scrutiny of those eco-attributes. Specifically, we assume that the cost associated with this is proportional to 1 – proportion of products being badged, i.e., cost for badges = m_{cb} x (1 – badge threshold), where

[27] The commissions for most categories have remained 15% on Amazon, 10% on eBay, and 5% on Tmall for years.

m_{cb} is a coefficient. For the two-seller game, various badge thresholds lead to three situations. When $I_0 = 0$, neither of the two products is badged, representing a baseline case in which the unified badge policy has not yet been implemented by the marketplace; when $I_0 = 0.5$, only one of the products is badged; and when, $I_0 = 1$, both products are badged. These three situations are extreme cases, as in reality, there are many products, and the badge proportion can take continuous values; it also cannot be 0 or 1. For simplicity, we consider only these three cases, given the two competing products. We assume that the feature of being badged is not incorporated into the recommendation system of the marketplace. Consistently, our empirical analysis of the Amazon search data did not find evidence that a badge had an effect on search rank, possibly because the green badge policy was a new practice (see Web Appendix D).

Consumers

In line with previous literature, we assume that consumers are heterogeneous in their horizontal preferences toward products A and B (Singh and Vives 1984). We model this heterogeneity as two segments: the green consumer segment of size $k \in (0, 1)$ and the non-green consumer segment of size $1 - k \in (0, 1)$. Both segments are aware of both products, A and B.

Segment 1: Green consumers. These consumers have a positive preference for the CPF badge. Specifically, consumer i's preference for the two products can be represented by the following utility function (Singh and Vives 1984):

$$u_{i,G} = (\alpha_{Ai} + \rho_i G_A) q_{iA} + (\alpha_{Bi} + \rho_i G_B) q_{iB} - \frac{\beta}{2}(q_{iA}^2 + q_{iB}^2 + 2z q_{iA} q_{iB}) - p_A q_{iA} - p_B q_{iB}. \quad (1)$$

In Equation (1), q_{ij} is consumer i's consumption quantity of product j, while p_j is the price of product j set by the seller; $G_j := \mathbb{1}(f_j \geq I_0) \in \{0, 1\}$ is the dummy variable indicating whether product $j \in \{A, B\}$ is badged, with f_j being the sustainability level of product j and I_0 being the badging threshold set by the marketplace. Coefficient α_{ji} captures consumer i's heterogeneous

marginal utility from consuming product j ($j \in \{A, B\}$) based on its features, excluding sustainability-related features, and follows a uniform distribution on $[\alpha_{j0} - c, \alpha_{j0} + c]$; ρ_i captures consumer i's heterogeneous marginal utility on the sustainability level from consuming either product, and we assume that ρ_i follows a uniform distribution on $[\rho_0 - m, \rho_0 + m]$. We assume that $\rho_0 \geq m$ and $\alpha_{j0} \gg c$ so that consumer i's marginal utility for product features unrelated to the sustainability of product j, α_{ji}, is always greater than 0, and consumer i's marginal utility for product features related to the sustainability of product j is always greater than or equal to 0, as some consumers might have a neutral attitude (Olsen, Slotegraaf, and Chandukala 2014). We use α_{j0} and ρ_0 to denote the baseline utility for product j's non-sustainability-related features and sustainability-related features, respectively. Additionally, $\beta > 0$ captures consumer i's extent of diminishing marginal utility from further consumption, and $0 < z < 1$ measures the level of substitutability or similarity between A and B for a given consumer (i.e., how a product's price will affect their purchase quantity of the other product); the larger parameter z is, the higher the similarity between two products.

To explore the effect of badge adoption on demand, we assume that the consumers in the green segment maximize their utility by simultaneously deciding the purchase quantities of the two products. We look at the first-order conditions (FOCs)

$$\frac{\partial u_{i,G}}{\partial q_{iA}} = \alpha_{Ai} + \rho_i G_A - \beta(q_{iA} + z q_{iB}) - p_A = 0, \tag{2}$$

$$\frac{\partial u_{i,G}}{\partial q_{iB}} = \alpha_{Bi} + \rho_i G_B - \beta(q_{iB} + z q_{iA}) - p_B = 0. \tag{3}$$

It follows that their purchase quantities are

$$q_{iA,G} = [\alpha_{Ai} + \rho_i G_A - p_A - z(\alpha_{Bi} + \rho_i G_B - p_B)]/[\beta(1 - z^2)], \tag{4}$$

$$q_{iB,G} = [\alpha_{Bi} + \rho_i G_B - p_B - z(\alpha_{Ai} + \rho_i G_A - p_A)]/[\beta(1 - z^2)]. \tag{5}$$

Segment 2: Non-green consumers. The consumers in this segment have negative preferences toward the CPF badge as they associate it with reduced product quality. Specifically, consumer i's preference can be represented by the following utility function (Singh and Vives 1984):

$$u_{i,N} = (\alpha_{Ai} - \rho_i G_A)q_{iA} + (\alpha_{Bi} - \rho_i G_B)q_{iB} - \frac{\beta}{2}(q_{iA}^2 + q_{iB}^2 + 2zq_{iA}q_{iB}) - p_A q_{iA} - p_B q_{iB}. \quad (6)$$

where q_{ij} is consumer i's consumption quantity of product j, while p_j is the price of product j. For coefficients, α_{ji} captures consumers' heterogeneous marginal utility from consuming product j ($j \in \{A, B\}$) based on its non-sustainability-related features, and follows a uniform distribution on $[\alpha_{j0} - c, \alpha_{j0} + c]$. We further assume that $\alpha_{j0} \gg c$ so that the marginal utility of consumer i for the other features of product j, excluding sustainability features, α_{ij}, remains positive. Similarly, ρ_i captures consumer i's heterogeneous marginal utility on the sustainability level from consuming either product, and we assume that ρ_i follows a uniform distribution of $[\rho_0 - m, \rho_0 + m]$. We further assume that $\rho_0 \geq m$ so that consumer i's marginal utility for product features related to the sustainability of product j is always less than or equal to 0, as some consumers might have a neutral attitude. As shown in $(\alpha_{ji} - \rho_i G_j)$, $j \in \{A, B\}$, when product *j* is badged, non-green consumers experience reduced utility, aligning with the sustainability-liability effect (Luchs et al. 2010), in which green labels are sometimes interpreted as signals of lower quality. Additionally, $\beta > 0$ captures the consumer's extent of diminishing marginal utility from further consumption, and $0 < z < 1$ measures the level of substitutability between A and B. Consumers buy products to maximize their utility. The FOCs

$$\frac{\partial u_{i,N}}{\partial q_{iA}} = \alpha_{Ai} - \rho_i G_A - \beta(q_{iA} + zq_{iB}) - p_A = 0, \quad (7)$$

$$\frac{\partial u_{i,N}}{\partial q_{iB}} = \alpha_{Bi} - \rho_i G_B - \beta(q_{iB} + zq_{iA}) - p_B = 0, \quad (8)$$

give us the optimal quantities of both products as

$$q_{iA,N} = [\alpha_{Ai} - \rho_i G_A - p_A - z(\alpha_{Bi} - \rho_i G_B - p_B)]/[\beta(1 - z^2)], \quad (9)$$

$$q_{iB,N} = [\alpha_{Bi} - \rho_i G_B - p_B - z(\alpha_{Ai} - \rho_i G_A - p_A)]/[\beta(1 - z^2)]. \quad (10)$$

Game Structure

The game has three stages (see Figure 1). In the first stage, the marketplace sets the badge threshold for green products from three choices $I_0 \in \{0, 0.5, 1\}$. The threshold here represents the percentage of the products receiving the badge, such that when the badge threshold is 0, no product will be badged, and when it is 1, both products will be badged. In the second stage, the two sellers simultaneously set their prices, p_A and p_B. In the last stage, consumers make their purchase decisions, and profits are realized for the marketplace and the sellers.

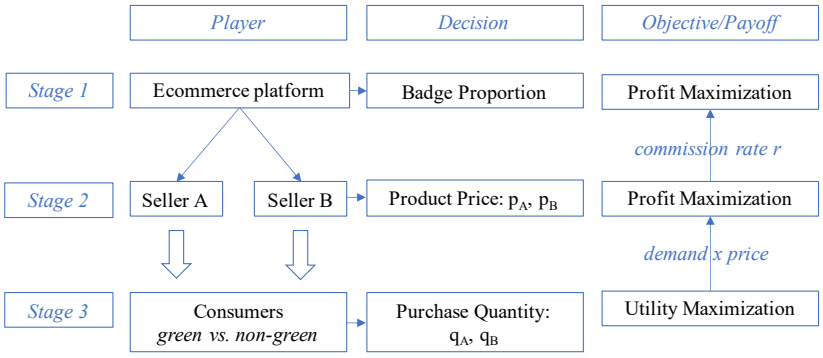


Figure 1. Three-Stage Game Structure

Equilibrium Outcome

This three-stage game has a few noteworthy equilibrium outcomes. First, under certain conditions, adopting a unified green badge leads to higher prices, increased demand, and reduced market concentration. This balance is achieved as long as the green segment of consumers with a positive attitude toward the badge is larger than the non-green segment of consumers with a negative attitude toward the badge. Second, an optimal balance occurs when the badging proportion is around 50%; both the marketplace and the sellers can maximize their profits compared to scenarios in which the badging proportion is either 0% or 100%. This optimal scenario arises when the baseline utility of non-sustainability-related features for the badged

product is at least twice that of the unbadged product. Additionally, the degree of product differentiation between the two sellers must be sufficiently large. Finally, the proportion of eco-conscious consumers with a positive perception of the badge should outweigh the proportion of non-green consumers who hold a negative perception.

We proceed with our analysis as follows. First, we assume that the marketplace endogenously decides the badge threshold and analyze the game with backward induction, starting from the seller's profit maximization. Given the badge threshold, the demand for product j (j ∈ {A, B}) is

$$D_j = \int_{\rho_0-m}^{\rho_0+m} \int_{\alpha_{B0}-c}^{\alpha_{B0}+c} \int_{\alpha_{A0}-c}^{\alpha_{A0}+c} [k \cdot q_{ij,G} + (1-k) \cdot q_{ij,N}] dF_{\alpha_{Ai}}(\alpha_{Ai})\, dF_{\alpha_{Bi}}(\alpha_{Bi})\, dF_{\rho_i}(\rho_i). \quad (11)$$

Next, we derive the profit-maximizing price and demand of the two sellers for the three choices of the badging threshold set by the marketplace in the first stage.

Situation 1: Badge proportion = 0. This baseline situation, in which no product is badged, is the same as the situation in which the marketplace has not implemented the unified badge policy. Therefore, $G_j = 0$, and the demand for both the green and non-green segments are the same. The demand for product A is

$$D_A^{(1)} = \frac{1}{\beta(1-z^2)} [\alpha_{A0} - z\alpha_0 - p_A + zp_B], \quad (12)$$

and the demand for product B is

$$D_B^{(1)} = \frac{1}{\beta(1-z^2)} [\alpha_{B0} - z\alpha_{A0} - p_B + zp_A]. \quad (13)$$

Thus, the profit derived from selling product A is

$$\Pi_A^{(1)} = (1-r)p_A \cdot \frac{1}{\beta(1-z^2)} [\alpha_{A0} - z\alpha_{B0} - p_A + zp_B], \quad (14)$$

and the profit derived from selling product B is

$$\Pi_B^{(1)} = (1-r)p_B \cdot \frac{1}{\beta(1-z^2)} [\alpha_{B0} - z\alpha_{A0} - p_B + zp_A]. \quad (15)$$

When maximizing the profit functions of products A and B simultaneously, there is only one root when FOC = 0. Therefore, we have

$$p_A^{*,(1)} = \frac{\alpha_{A0} - z\alpha_{B0} + zp_B^{*,(1)}}{2}, p_B^{*,(1)} = \frac{\alpha_{B0} - z\alpha_{A0} + zp_A^{*,(1)}}{2}, \quad (16)$$

which gives us the final expression for optimal prices and demand, as summarized in Lemma 1.

Lemma 1. *When the badging threshold is 0, no product is badged, and the profit-maximizing subgame equilibrium has optimal prices as* $p_A^{*,(1)} = \frac{1}{4-z^2}[(2-z^2)\alpha_{A0} - z\alpha_{B0}]$ *and* $p_B^{*,(1)} = \frac{1}{4-z^2}[(2-z^2)\alpha_{B0} - z\alpha_{A0}]$*, and the demand for products A and B of this subgame perfect equilibrium is* $D_A^{*,(1)} = \frac{1}{\beta(1-z^2)(4-z^2)}[(2-z^2)\alpha_{A0} - z\alpha_{B0}]$ *and* $D_B^{*,(1)} = \frac{1}{\beta(1-z^2)(4-z^2)}[(2-z^2)\alpha_{B0} - z\alpha_{A0}]$, respectively.

To ensure nonnegative equilibrium prices, we imposed the following constraint:

$$\alpha_{A0} > \frac{z}{2-z^2}\alpha_{B0} > \left(\frac{z}{2-z^2}\right)^2 \alpha_{A0}, \quad (17)$$

which, given that we assume $0 < z < 1$, as it measures the level of substitutability between products A and B for a given consumer, such a condition is satisfied as long as $\alpha_{A0} > \alpha_{B0} > \frac{z}{2-z^2}\alpha_{A0}$.

Situation 2: Badge proportion = 0.5. This is an asymmetric structure in which only product A is badged, as $f_A > I_0$ while $f_B < I_0$. Thus, $G_A = 1$ while $G_B = 0$, and the utility functions of both consumer segments are different due to different consumer preferences for the badge. The demand for product A is

$$D_A^{(2)} = \frac{k}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} + \rho_0 - p_A + zp_B] + \frac{1-k}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} - \rho_0 - p_A + zp_B]$$

$$= \frac{1}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} + (2k-1)\rho_0 - p_A + zp_B], \quad (18)$$

whereas the demand for product B is

$$D_B^{(2)} = \frac{k}{\beta(1-z^2)}[\alpha_{B0} - z\alpha_{A0} - z\rho_0 - p_B + zp_A)] + \frac{1-k}{\beta(1-z^2)}[\alpha_{B0} - z\alpha_{A0} + z\rho_0 - p_B + zp_A)]$$

$$= \frac{1}{\beta(1-z^2)}[\alpha_{B0} - z\alpha_{A0} - z(2k-1)\rho_0 - p_B + zp_A], \quad (19)$$

Thus, the profit derived from selling product A is

$$\Pi_A^{(2)} = (1-r)p_A \cdot \frac{1}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} + (2k-1)\rho_0 - p_A + zp_B], \quad (20)$$

and the profit derived from selling product B is

$$\Pi_B^{(2)} = (1-r)p_B \cdot \frac{1}{\beta(1-z^2)}[\alpha_{B0} - z\alpha_{A0} - z(2k-1)\rho_0 - p_B + zp_A]. \quad (21)$$

We maximize the profit derived from selling products A and B simultaneously to obtain optimal prices,

$$p_A^{*,(2)} = \frac{\alpha_{A0} - z\alpha_{B0} + (2k-1)\rho_0 + zp_B^{*,(2)}}{2}, \; p_B^{*,(2)} = \frac{\alpha_{B0} - z\alpha_{A0} - z(2k-1)\rho_0 + zp_A^{*,(2)}}{2}, \quad (22)$$

which thus gives us the final expression for optimal prices and demand, as summarized in Lemma 2.

Lemma 2*. When the badging threshold is 0.5, only product A is badged, and the profit-maximizing subgame equilibrium has an optimal price as* $p_A^{*,(2)} = \frac{1}{4-z^2}[(2-z^2)\alpha_{A0} - z\alpha_{B0} + (2-z^2)(2k-1)\rho_0]$, $p_B^{*,(2)} = \frac{1}{4-z^2}[(2-z^2)\alpha_{B0} - z\alpha_{A0} - z(2k-1)\rho_0]$*. The demand for products A and B of this subgame perfect equilibrium are* $D_A^{*,(2)} = \frac{1}{\beta(1-z^2)(4-z^2)}[(2-z^2)\alpha_{A0} - z\alpha_{B0} + (2-z^2)(2k-1)\rho_0]$ *and* $D_B^{*,(2)} = \frac{1}{\beta(1-z^2)(4-z^2)}[(2-z^2)\alpha_{B0} - z\alpha_{A0} - z(2k-1)\rho_0]$.

Since we are interested in the effect of the CPF badge, we only looked at the change in price and demand of product A before and after receiving the badge. Based on Lemma 1 and Lemma 2, by comparing the price difference for product A before and after it is badged in situation 1 and situation 2,

$$p_A^{*,(2)} - p_A^{*,(1)} = \frac{(2-z^2)(2k-1)\rho_0}{4-z^2}, \tag{23}$$

we can see that the optimal prices for product A in situation 2 are larger than those in situation 1 and thus larger than 0, as long as there are more consumers from the green segment than from the non-green segment, i.e., k > 0.5. We can conclude that the price increase in situation 2 is due to the fact that consumers in the green segment with positive attitudes to the badge outnumber consumers from the non-green segment with negative attitudes, which can be understood as the price premium caused by green features among eco-conscious consumers.

If the demand for product A in situation 2 is greater than that in situation 1,

$$D_A^{*,(2)} - D_A^{*,(1)} = \frac{(2-z^2)(2k-1)\rho_0}{\beta(1-z^2)(4-z^2)} > 0, \tag{24}$$

which is true as long as k > 0.5 (given that we already assume 0 < z < 1), this means that the number of green consumers is greater than the number of non-green consumers. This can be attributed to the increased utility among the green segment of consumers for badged products outweighing the sum of decreased utility among the non-green segment of consumers and the decreased utility due to the price increase.

Proposition 1*. The price and demand for badged product A in situation 2 (one product badged) are larger than those in situation 1 when the product is not badged, as long as the number of green consumers is greater than the number of non-green consumers.*

The intuition for this condition is straightforward. Provided that consumers who perceive the unified green badge favorably outnumber those who view it negatively, overall demand will rise, despite a price premium. This is because the perceived utility of products with the badge increases among the majority, who are eco-conscious consumers.

If the demand and price for badged product A in situation 2 is smaller than that in situation 1, then we need to assume that the number of non-green consumers is greater than the number of green consumers, which means that k < 0.5.

Situation 3: Badge threshold = 1. In this situation, both products are badged (i.e., $G_A = 1$ and $G_B = 1$). The demand for product A is

$$D_A^{(3)} = \frac{1}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} + (2k-1)(1-z)\rho_0 - p_A + zp_B], \tag{25}$$

and the demand for product B is

$$D_B^{(3)} = \frac{1}{\beta(1-z^2)}[\alpha_{B0} - z\alpha_{A0} + (2k-1)(1-z)\rho_0 - p_B + zp_A]. \tag{26}$$

Thus, the profit derived from selling product A is

$$\Pi_A^{(3)} = (1-r)p_A \cdot \frac{1}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} + (2k-1)(1-z)\rho_0 - p_A + zp_B], \tag{27}$$

and the profit derived from selling product B is

$$\Pi_B^{(3)} = (1-r)p_B \cdot \frac{1}{\beta(1-z^2)}[\alpha_{B0} - z\alpha_{A0} + (2k-1)(1-z)\rho_0 - p_B + zp_A]. \tag{28}$$

Profit maximization implies that the optimal prices of products A and B are

$$p_A^{*,(3)} = \frac{\alpha_{A0} - z\alpha_{B0} + (2k-1)(1-z)\rho_0 + zp_B^{*,(3)}}{2} \text{ and } p_B^{*,(3)} = \frac{\alpha_{B0} - z\alpha_{A0} + (2k-1)(1-z)\rho_0 + zp_A^{*,(3)}}{2}, \tag{29}$$

which thus gives us the final expression for optimal prices and demand, as summarized in Lemma 3.

Lemma 3. *When the badging threshold is 1, both products are badged, and the profit-maximizing subgame equilibrium has an optimal price* $p_A^{*,(3)} = \frac{1}{4-z^2}[(2-z^2)\alpha_{A0} - z\alpha_{B0} + (2k-1)(z+2)(1-z)\rho_0]$, $p_B^{*,(3)} = \frac{1}{4-z^2}[(2-z^2)\alpha_{B0} - z\alpha_{A0} + (2k-1)(z+2)(1-z)\rho_0]$. *The demand for products A and B are* $D_A^{*,(3)} = \frac{1}{\beta(1-z^2)(4-z^2)}[(2-z^2)\alpha_{A0} - z\alpha_{B0} +$

$(2k-1)(z+2)(1-z)\rho_0]$ and $D_B^{*,(3)} = \frac{1}{\beta(1-z^2)(4-z^2)}[(2-z^2)\alpha_{B0} - z\alpha_{A0} + (2k-1)(z+2)(1-z)\rho_0]$.

Given that $0 < z < 1$, as it measures the level of substitutability between products A and B for a given consumer, the prices are positive $\alpha_{A0} > \alpha_{B0} > \frac{z}{2-z^2}\alpha_{A0}$, as specified in Equation (17). Based on Lemmas 1 and 3, the sub-equilibrium prices in situation 3 for both badged products, A and B, are larger than those in situation 1 where they are not badged, as long as $k > 0.5$, which means that the number of green consumers is greater than the number of non-green consumers. Similarly, if we want the demand for products A and B in situation 3 to be at least larger than that in situation 1, $k > 0.5$ is the sufficient and necessary condition to be met.

Proposition 2. *The sufficient and necessary condition that prices and demand in situation 3 (both products badged) are larger than those in situation 1 (no product badged) is that the number of green consumers is greater than the number of non-green consumers.*

The reasoning behind this proposition is straightforward. If the number of consumers who view the unified green badge favorably exceeds those with a negative perception, the overall demand for badged products will increase, even with a price premium. This increase in demand is attributed to the heightened perceived utility among eco-conscious consumers, who form the majority. Additionally, the utility gains for this consumer segment outweigh any decrease in utility due to the price premium and the negative views of non-green consumers about the badge.

Finally, if the total profit for both products A and B on behalf of the marketplace in situation 2 is larger than that in situation 3, such that the medium-level badging threshold is the optimal choice for the marketplace, then the following conditions hold. Based on Lemma 2 and 3, when comparing prices:

$$\xi_A^{23} = p_A^{*,(2)} - p_A^{*,(3)} = \frac{1}{4-z^2}z(2k-1)\rho_0,$$

$$\xi_B^{23} = p_B^{*,(2)} - p_B^{*,(3)} = -\frac{1}{4-z^2}(2-z^2)(2k-1)\rho_0 = -\frac{2-z^2}{z}\xi_A^{23}. \quad (30)$$

Similarly, when comparing demands for both products,

$$D_A^{*,(2)} - D_A^{*,(3)} = \frac{1}{\beta(1-z^2)(4-z^2)} z(2k-1)\rho_0 = \frac{1}{\beta(1-z^2)}\xi_A^{23},$$

$$D_B^{*,(2)} - D_B^{*,(3)} = -\frac{1}{\beta(1-z^2)(4-z^2)}(2k-1)(2-z^2)\rho_0 = \frac{1}{\beta(1-z^2)}\xi_B^{23}. \quad (31)$$

If the total profit obtained by the marketplace after the commission rate for situation 2 is $\Pi_M^{(2)}$, then we have

$$\Pi_M^{(2)} = r \cdot [p_A^{*,(2)} \cdot D_A^{*,(2)} + p_B^{*,(2)} \cdot D_B^{*,(2)}] - 0.5m_{cb}, \quad (32)$$

$$\Pi_M^{(3)} = r \cdot [p_A^{*,(3)} \cdot D_A^{*,(3)} + p_B^{*,(3)} \cdot D_B^{*,(3)}], \quad (33)$$

Thus, when comparing the profits, we have

$$\Pi_M^{(2)} - \Pi_M^{(3)} = \frac{r}{\beta(1-z^2)(4-z^2)z}\xi_A^{23} \cdot [4z(2-z^2)\alpha_{A0} - 2z(z^4-2z^2+z+4)\alpha_{B0} + (-z^4 - 4z^3+3z^2+8z-4)(2k-1)\rho_0] + 0.5m_{cb}. \quad (34)$$

We can observe that a sufficient condition for $\Pi_M^{(2)} - \Pi_M^{(3)} > 0$ is

$$4z(2-z^2)\alpha_{A0} - 2z(z^4-2z^2+z+4)\alpha_{B0} > 0, \quad (35)$$

$$(-z^4-4z^3+3z^2+8z-4)(2k-1)\rho_0 > 0, \quad (36)$$

which can be satisfied as long as

$$\frac{\alpha_{A0}}{\alpha_{B0}} > 2 > \frac{z^4+z}{4-2z^2} + 1, -z^4-4z^3+3z^2+8z-4 > 0, \text{ and } k > 0.5,$$

which can be satisfied when k > 0.5 and z ∈ [0.48, 1].

Finally, because prices and quantities in situation 3 exceed those in situation 1 whenever k > 0.5, and because the marketplace bears the extra audit cost $0.5\,m_{cb}$ only in situation 1, we have $\Pi_M^{(2)} > \Pi_M^{(3)}$ and $\Pi_M^{(2)} > \Pi_M^{(1)}$. Hence, the following proposition holds.

Proposition 3. *Profit maximization for the marketplace and both sellers can be achieved*[28] *at the medium level of the badging threshold when the following three sufficient conditions hold: (1)* $\alpha_{A0} > 2\alpha_{B0}$*, (2)* $z \in [0.48, 1]$*, and (3)* $k > 0.5$*; that is, the baseline utility for non-sustainability-related features of the badged product is at least double that of the unbadged product, the product differentiation between two sellers is sufficiently high, and the green segment of consumers outnumbers the non-green segment.*

For example, when k = 0.6, z = 0.8, $\alpha_{A0} = 30$, and $\alpha_{B0} = 8$, all of the sufficient conditions are met, so profit is maximized for both the marketplace and the two sellers at the medium badge threshold.

Comparisons of market concentration. Finally, we looked at market concentration under three situations. To capture market concentration, we calculated the Herfindahl–Hirschman index (HHI), which is the sum of the squares of the market shares of all firms operating in a particular market; this indicator has been widely adopted in past research (Kelly 1981). The HHI gives higher weights to larger firms, with higher values indicating greater market concentration, and is calculated as follows:

$$\text{HHI} = \sum_{i=1}^{n} s_i^2, \tag{37}$$

where n is the number of firms in the market and s_i is the market share of firm i. The market share is

$$MS_A = \frac{D_A^*}{D_A^* + D_B^*}, MS_B = \frac{D_B^*}{D_A^* + D_B^*}, MS_A + MS_B = 1, \tag{38}$$

based on which we can calculate the HHI as

$$HHI^* = (MS_A)^2 + (MS_B)^2 = (MS_A)^2 + (1 - MS_A)^2 \geq 0.25, \tag{39}$$

[28] Since we are looking for a sufficient condition of increased profit for both the platform and the sellers, we compare price and demand separately so that the highest profit is achieved when both price and demand are highest among the three situations.

which takes the minimum when $MS_A = MS_B$, or, in other words, when $D_A^* = D_B^*$. Based on the U-shape curve between HHI and MS_A (Figure 2), we can now derive that the closer MS_A is to 0.5, the smaller the HHI, and the lower the market concentration.

By looking at the subgame perfect equilibrium under three situations, we have

$$|MS_A^{(1)} - \frac{1}{2}| = \frac{|D_A^{*,(1)} - D_B^{*,(1)}|}{2(D_A^{*,(1)} + D_B^{*,(1)})} = \frac{(2-z^2+z)|\alpha_{A0} - \alpha_{B0}|}{2(2-z^2-z)(\alpha_{A0} + \alpha_{B0})}, \quad (40)$$

$$|MS_A^{(2)} - \frac{1}{2}| = \frac{|D_A^{*,(2)} - D_B^{*,(2)}|}{2(D_A^{*,(2)} + D_B^{*,(2)})} = \frac{(2-z^2+z)|\alpha_{A0} - \alpha_{B0} + (2k-1)\rho_0|}{2(2-z^2-z)(\alpha_{A0} + \alpha_{B0} + (k-0.5)\rho_0)}, \quad (41)$$

$$|MS_A^{(3)} - \frac{1}{2}| = \frac{|D_A^{*,(3)} - D_B^{*,(3)}|}{2(D_A^{*,(3)} + D_B^{*,(3)})} = \frac{(2-z^2+z)|\alpha_{A0} - \alpha_{B0}|}{2(2-z^2-z)(\alpha_{A0} + \alpha_{B0} + (2k-1)\rho_0)}. \quad (42)$$

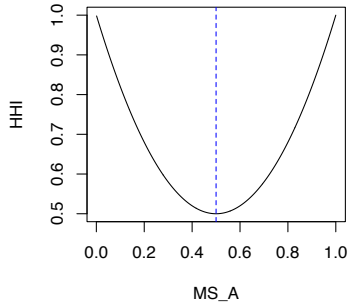


Figure 2. U-shaped Relationship Between Market Share of Product A (MS_A) and Market Concentration (Herfindahl–Hirschman Index, HHI)

When we compare the three situations, we can see that as long as k > 0.5, $|MS_A^{(3)} - \frac{1}{2}| < |MS_A^{(1)} - \frac{1}{2}|$ and $|MS_A^{(3)} - \frac{1}{2}| < |MS_A^{(2)} - \frac{1}{2}|$, which gives us the following proposition.

Proposition 4. *Comparing situation 3 (both products badged) with situation 1 (no product badged) and situation 2, to achieve higher price and demand and lower market concentration in situation 3, one sufficient condition is that the green segment is larger than the non-green segment of consumers.*

To summarize, if the equilibrium represents higher price and demand and lower market concentration when the unified green badge policy is implemented than when no product is badged, then the number of green consumers with positive attitudes toward the badge is greater than the number of non-green consumers with negative attitudes toward the badge.

Discussions on Model Assumptions

We now turn to a detailed discussion of the model's underlying assumptions. First, we posit two distinct consumer segments: "green" buyers who derive additional utility from badged products and "non-green" buyers who hold a negative bias. Were there only one consumer type, the trade-off between increased demand from badge-favoring shoppers and decreased demand from badge-averse shoppers would vanish, fundamentally altering our equilibrium predictions.

Second, we treat Amazon's CPF badge allocation as a bidirectional, selective process—only products meeting stringent environmental criteria receive the badge (Web Appendix A). If every green-eligible product automatically obtained the badge, tag adoption would exclusively depend on the green-consumer majority, leading to uniformly higher prices and demand.

Third, we assume that all consumers, regardless of segment, consider both badged and unbadged options but that the badge modifies their utility and thus their purchase quantities. This generalization preserves the consistency of our original model's conclusions even when only green consumers are no longer assumed to buy only badged products; the badge continues to shift choices in predictable ways.

Fourth, we assume that the CPF badge does not affect Amazon's search rankings. If the badge actually pushed products higher in search results, some of the demand boost we observed would come from this added visibility rather than from the badge's sustainability signal alone.

Finally, we assume that consumers' marginal utility for product features unrelated to sustainability is not correlated with the likelihood of obtaining a badge. We make this assumption based on our interview with Amazon—whether a product can have a CPF badge is based solely on its sustainability features. While we cannot test this assumption directly, our Amazon data show patterns—higher demand, higher prices, and lower market concentration for badged products—that align with our model's theoretical predictions under a moderate badging strategy and a green-majority consumer base. These results remain robust even if the badge also influences utility from non-sustainability features, provided that influence is the same for both products. In addition, if there are three competing sellers, with two of them selling green products and incurring the cost of maintaining the green status and one of them being non-green, the CPF badge can still lead to increased demand and price (see detailed discussion in Web Appendix A).

DATA

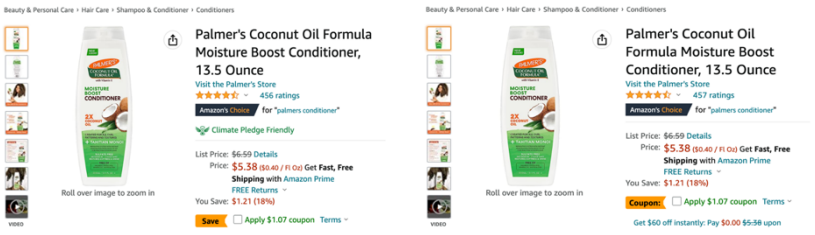
To explore the causal impact of unified green badge adoption on product demand, price, and market concentration, and to assess the potential for a win-win-win outcome as predicted by our theoretical model, we collected 6.5 months of data from Amazon on 6,606 distinct products across 8 categories and 20 subcategories. Of these products, 35.4% had been badged with the CPF label at least once.

Price and Demand

We collected daily data on 14,000 products across eight categories with a high percentage of green products on Amazon from March 1, 2023 to September 15, 2023. These categories were beauty and personal care, health and household, grocery and gourmet food, clothing, shoes and jewelry, sports and outdoors, office products, electronics, and toys and games. After removing

incomplete observations and products with fewer than 135 days of data (out of 193 days), we restricted our sample to 6,606 distinct products. Sales rank, a widely used metric for quantifying demand on Amazon, was our primary measure, with lower ranks indicating higher sales. We identified products using their Amazon Standard Identification Number (ASIN), a 10-character alphanumeric identifier.[29] Table W1 in Web Appendix B details the range of products by type and subtype and the presence of green certifications.

For each product, we collected the following metrics daily: price, rating count, mean positive scores of reviews, and mean rating count. In addition, we collected the following metrics weekly: mean compound valence score; length, richness, and readability of product description; and colorfulness, brightness, symmetry, average face count, technical quality score, and aesthetics quality score of product images. During the period of our data collection, we observed instances of treatment reversal, where products experienced changes in their badge status. Specifically, some products initially did not have a badge but were later awarded one, while others initially had a badge but had it removed. This dynamic nature of badge allocation provided a unique opportunity to observe the effects of these changes. For example, the product shown in Figures 3 (a) and (b) initially possessed a badge but had it removed on May 18, 2023. Web Appendix B presents the history of CPF badge status for a random sample.



(a) Before the badge was removed (Mar 10) (b) After the badge was removed (Mar 11)

[29] See detailed examples and introduction at https://developer.amazon.com/docs/mobile-associates/mas-finding-product-id.html.

Figure 3. Example of Product Before Badge Removal (Left) and After Badge Removal (Right)

Product Images

We collected weekly product images from March 1, 2023 to September 15, 2023 on the 6,606 distinct products in our demand dataset. We used computer vision techniques and deep learning models to extract interpretable product features from unstructured image data. The list of features we extracted comprises those that have proven critical for consumer decisions: (1) colorfulness, (2) brightness, (3) visual balance, (4) visual complexity, (5) image quality, and (6) face count.

Colorfulness measures how chromatic (any color except white, gray, or black) the perceived color of an area appears to be (Fairchild 2013). The general *brightness* of an image is also important; ample illumination is crucial in rendering image content clear to viewers, as the information within images is communicated via pixel brightness (Gorn et al. 1997). *Visual balance* can be understood as the symmetry of an image's visual elements (in this case, intensity and color). Visually balanced real estate images give viewers a feeling of order and tidiness, minimizing the cognitive demand required to process the images (Kreitler and Kreitler 1972). *Visual complexity* can impact both liking and usability in advertisements. We used edge density as a proxy for visual complexity (Pieters, Wedel, and Batra 2010). *Image quality* can have a significant impact on consumer decisions and product demand, especially in online transactions (Zhang et al. 2022). We used the Neural IMage Assessment (NIMA) to assess the image quality of product images (Talebi and Milanfar 2018; Ceylan, Diehl, and Proserpio 2023). *Face count* can be understood as the approximate number of human models employed by the brand in the advertisement. Past literature has shown the significant impact that a human model, especially an attractive one, has on sales in a product advertisement (Baker and Churchill 1977).

Product Descriptions and Customer Reviews

We also collected weekly product descriptions from March 1, 2023 to September 15, 2023 on the 6,606 distinct products in the demand dataset. We used advanced natural language processing techniques and deep learning models to extract interpretable product features from unstructured text data. The list of features we extracted comprised (1) valence, (2) richness, and (3) readability. Note that we tracked only the reviews' characteristics, not the reviewers' characteristics.

Valence refers to the emotion conveyed in the product description, indicating whether the language evokes positive, negative, or neutral emotions among consumers. *Richness* refers to lexical richness and encompasses the breadth and diversity of vocabulary within a given text. This aspect of language can be a significant indicator of various factors, including writing quality and vocabulary knowledge (McCarthy and Jarvis 2007). Past research has shown that more readable descriptions are more memorable (Reczek et al. 2018). We used the Flesch Reading Ease Score test to assess each text's *readability* by analyzing word, syllable, and sentence counts.

Additionally, we collected all customer reviews listed on the "See more reviews" page as of September 16, 2023. Since Amazon frequently detects and deletes fake reviews (He, Hollenbeck, and Proserpio 2022), we did not rely solely on the rating count and score displayed on the product detail page. Instead, we calculated the newly added and total review counts and computed the average rating score based on authentic (non-deleted) reviews. We also performed sentiment analysis on customer reviews using Vader's sentiment analyzer (Hutto and Gilbert 2014), as previous research has shown that review sentiment impacts demand (Liu, Lee, and

Srinivasan 2019). In addition, we counted word occurrences related to sustainability[30] as a proxy for the extent to which the review focused on sustainability; we also counted word occurrences pertaining to packaging and return[31] as a proxy for the extent to which the review was written for the delivery process.

Multimodal Vector Representations

To eliminate concerns about the inadequate selection of image and description features, as introduced in the Product Images and Product Descriptions and Customer Reviews subsections, we alternatively extracted non-interpretable vector representations from a state-of-the-art multimodal machine learning model, Contrastive Language–Image Pre-training (CLIP; Radford et al. 2021). We chose CLIP mainly because its model demonstrates enhanced flexibility and generality compared to conventional ImageNet models. CLIP scales a straightforward pre-training task to attain competitive zero-shot performance across diverse image classification datasets.

Figure 4 shows the three steps involved in extracting the image and description vector representations. First, we used the NLTK summarizer to shorten the long descriptions, given that the CLIP model has a length restriction of 77 characters. Second, we input the image into the image feature encoder and the summarized description into the text feature encoder; both encoders had a backbone structure of ViT-B/32 Transformer architecture (Radford et al. 2021). Third, after obtaining a 512-dimensional vector representation from the image feature encoder and a 512-dimensional vector representation from the text feature encoder for all images and descriptions, we used principal component analysis to conduct dimension reduction for the

[30] The set of strings we used for calculating occurrences of sustainability-related topics included *eco*, *climate*, *environment*, *green*, *renew*, *preserv*, *endur*, *earth*, *bio*, and *sustain*.

[31] The set of strings we used for calculating occurrences on packaging-and-return related topics included *packag*, *return*, *parcel*, *box*, *empty*, *ship*, and *deliver*.

images and descriptions. We obtained a 10-dimensional vector representation for each product image and a 10-dimensional vector representation for each (summarized) product description.

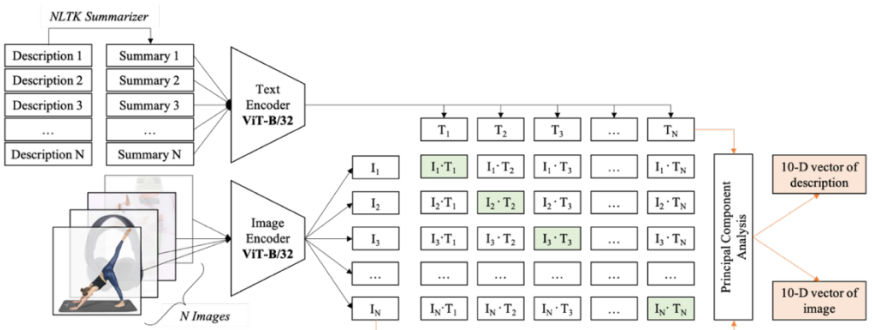


Figure 4. CLIP Framework for Extracting Image and Textual Vector Representations

EMPIRICAL METHODS

We employed the interactive fixed effects counterfactual (IFEct) estimator (Liu, Wang, and Xu 2022) to infer the causal impact of the CPF badge on a set of critical business outcomes: product demand, price, and market concentration. The IFEct approach fulfills our research objective for two reasons. First, it allows for time-varying covariates, which was necessary because there were many time-varying features to be controlled, including price, rating score, product image, and description features. Second, it allows for treatment reversal, which appeared in our data; some products did not have a badge initially but were later awarded one, while others that initially had a badge had it removed.

The IFEct Framework for Identifying the Causal Effects of the CPF Badge

Consider a balanced sample that consists of N units observed over T periods. Here D_{it} denotes the dummy variable, with 1 representing being treated and 0 otherwise; Y_{it} is the outcome variable, with $Y_{it}(0)$ denoting the outcome when it is not treated and $Y_{it}(1)$ when it is treated; $\boldsymbol{X}_{it}$ is a vector of exogenous covariates; $\boldsymbol{U}_{it}$ is a vector of unobservable covariates; and ε_{it} is the idiosyncratic error term.

We were interested in quantifying the average treatment effect on the treated units (treatment status changed at least once during the observed time window):

$$ATT = \mathbb{E}[\delta_{it}|D_{it} = 1, C_i = 1], \tag{43}$$

where $\delta_{it} = Y_{it}(1) - Y_{it}(0)$; $C_i = 1$ if unit i had its treatment status changed at least once, i.e., $\exists t, s \in 1,2, \dots, T \text{such that} D_{it} = 1, D_{is} = 0$; otherwise, $C_i = 0$.

We followed Liu, Wang, and Xu (2022) and made three identification assumptions: no temporal or spatial interaction, strict exogeneity, and the existence of a low dimensional decomposition. Detailed discussions are provided in Web Appendix C.

Augmented Factors and Relaxation on Exogeneity

When unobserved time-varying confounders exist, there are concerns about the endogeneity issue, and the exogeneity assumption will not hold. Some authors have proposed using factor-augmented models to relax the strict exogeneity assumption (Bai 2009; Xu 2017). Among these, IFEct models the response surface of untreated potential outcomes using a factor-augmented model. We assume that

$$Y_{it}(0) = \boldsymbol{X}'_{it}\beta + \alpha_i + \tau_t + \lambda'_i f_t + \varepsilon_{it}, \tag{44}$$

where $f_t = [f_{1t}, \dots, f_{rt}]' \in \mathbb{R}^{r\times 1}$ is a vector of unobserved common factors, and $\lambda_i = [\lambda_{i1}, \dots, \lambda_{ir}]' \in \mathbb{R}^{r\times 1}$ is a vector of unknown factor loadings. We assume that this factor component takes a linear, additive form $\lambda'_i f_t = \sum_{k=1}^{r} \lambda_{ik} f_{kt}$. In general, as long as an unobserved random variable can be decomposed into a multiplicative form, it can be absorbed by $\lambda'_i f_t$. Here $\boldsymbol{X}'_{it}$ is a vector of covariates; α_i is unit fixed effect, while τ_t is time fixed effect; and ε_{it} is a matrix of idiosyncratic errors (see Figure 5). Detailed steps in the estimation of the IFEct estimator are provided in Web Appendix C.

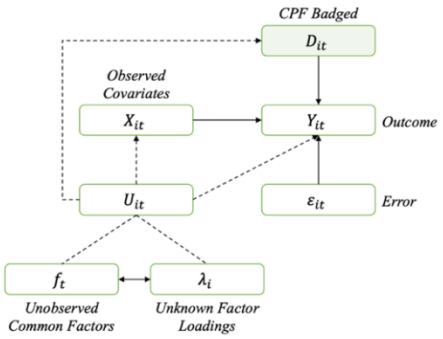


Figure 5. Directed Graphical Model Illustration of the IFEct Estimator

Finally, we thoroughly assessed potential biases and confounders, including a pre-trend analysis, placebo tests, and evaluations for any carryover effects, bolstering the credibility of our findings. We also investigated the CPF badge's influence on product search rankings on the platform and did not find evidence that it altered them. Details are provided in Web Appendix D.

EMPIRICAL RESULTS

We present our findings on the causal effects of adopting the CPF badge on product demand, pricing strategies, and market concentration. This analysis tests our three main hypotheses.

CPF Badge Adoption Leads to Increased Demand

We first explored the causal impact of CPF on demand, measured as log(sales rank) (He, Hollenbeck, and Proserpio 2022; Park, Xie, and Xie 2023). We ran two models with different sets of covariates. Model (1) included only the price and rating count. Model (2) added price, rating count, mean positive review scores, mean rating count, mean compound valence score, and various features of product descriptions (length, richness, and readability) and images (colorfulness, brightness, symmetry, average face count, technical quality, and aesthetic quality). The summary statistics are shown in Table 1.

Table 1. Summary Statistics of Key Variables

VARIABLE	MEAN	MEDIAN	SD	MIN	MAX

Log sales rank	9.98	10.21	2.07	0.41	16.19
Log price	3.09	3.00	0.81	0.69	6.17
Log review #	4.14	4.28	1.25	0.69	6.80
Description compound	0.92	0.98	0.20	−0.81	1.00
Description length	705.38	683.00	358.83	9.00	2,615.00
Description richness	0.76	0.75	0.11	0.40	1.00
Description readability	48.83	54.29	27.08	−118.00	120.21
Colorfulness	110.92	107.64	41.42	8.79	274.54
Brightness	0.55	0.55	0.14	0.11	0.97
Symmetry	89,007.89	85,183.00	20,050.14	37,678.78	31,4528.76
Face #	0.31	0.50	0.24	0.10	2.29
Edge density	0.58	0.50	0.41	0.06	9.51
Image technical score	5.23	5.22	0.32	4.24	6.25
Image aesthetics score	9.98	10.21	2.07	0.41	16.19

We used four augmented factors to minimize root mean square error (RMSE). The average treatment effect on the treated (ATT) and estimated coefficients for each covariate are shown in Table 2, with time variation in the average treatment effect for both models illustrated in Figure 6.

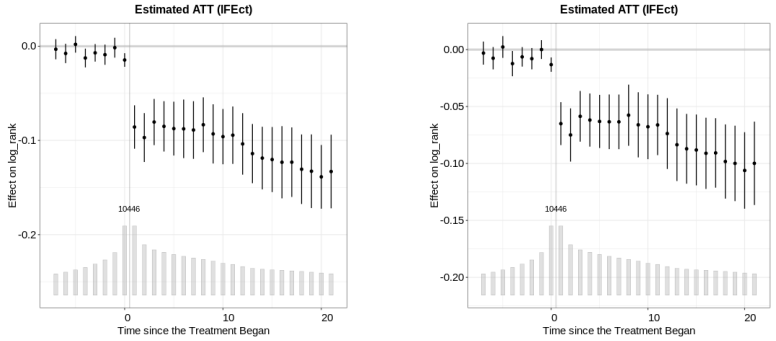
We observed a significant negative impact of adopting the CPF badge on sales rank, indicating a positive effect of this badge on sales volume. This result remained consistent even after controlling for various product features in Model (2). However, the effect size was smaller in Model (2) than in Model (1), suggesting that product images and descriptions may significantly influence how the CPF badge affects demand. Diagnostics verifying the robustness of Model (2)—including tests for pre-trend, placebo, and carryover effects—are reported in Web Appendix D. As an additional check, we found minimal overlap between the CPF badge and other Amazon badges. The correlation between CPF status and the “Best Seller” badge was very small (–0.036), indicating that CPF-badged items almost never overlapped with best sellers, so the post-badge demand lift was unlikely to stem from best seller status. The proportion of

products with the "Best Seller" badge was below 5% for both badged (2.22%) and non-badged products (0.94%). Likewise, the association between CPF status and "Limited-Time Deal" status was modest (0.103). The proportion of products with "Limited-Time Deal" status was around 80% for both badged (79.40%) and non-badged products (79.70%).

Table 2. Estimation of Causal Impact of CPF Badge on Demand (DV: Log Sales Rank, lower = higher demand)

DV: Log (sales rank)	**Model (1)**		**Model (2)**	
	ATT	p	ATT	p
Treated observations equally weighted	−0.1394	0.0000	−0.1041	0.0000
Treated units equally weighted	−0.1317	0.0000	−0.0974	0.0000
	$\hat{\beta}$	p	$\hat{\beta}$	p
Log (price + 1)	0.1060	0.0000	0.1279	0.0000
Log (rate # + 1)	−0.4897	0.0000	−0.5219	0.0000
Mean review positivity			0.4311	0.0872
Mean rating score			0.0479	0.0769
Log (sustainability topic + 1)			0.2781	0.0000
Log (packaging topic + 1)			0.0058	0.8441
Description valence			−0.1099	0.0877
Log (description length + 1)			−0.1174	0.0013
Description richness			−0.3331	0.0777
Description readability			−0.0020	0.0112
Colorfulness			0.0002	0.7816
Brightness			0.1305	0.4464
Symmetry			0.0000	0.4629
Face #			−0.1851	0.0679
Image technical score			0.0851	0.1295
Image aesthetic score			0.0304	0.7494
RMSE	0.3440		0.3461	
Obs. #	1,274,958		1,274,958	

Notes. For identification purposes, units whose number of untreated periods was less than 5 were dropped automatically.



(a) ATT in Model (1) (b) ATT in Model (2)

Figure 6. ATT of Having the CPF Badge on Demand from Model (1) [a] and Model (2) [b]

Notes. The gray bar denotes the number of units at the *t* period after treatment. DV is sales rank; higher rank indicates lower demand.

CPF Badge Adoption Leads to Increased Price

As shown in Table 3, we estimated two models to explore the causal impact of CPF on price. Model (3) included only sales rank and rating count. Model (4) added sales rank, rating count, review count, mean positive review scores, mean compound valence score, and various features of product descriptions (length, richness, and readability) and images (colorfulness, brightness, symmetry, average face count, technical quality, and aesthetic quality). We used four augmented factors to minimize the RMSE. The average treatment effect and estimated coefficients for each covariate are shown in Table 3, with time variation in the ATTs plotted in Figure 7.

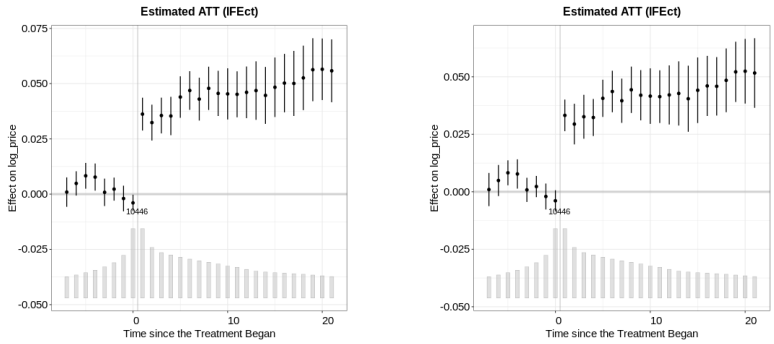
We observed a positive and significant impact of adopting the CPF badge on price, consistent even when controlling for various product features in Model (4). This aligns with previous findings that consumers are willing to pay a premium for green products (Tully and Winer 2014). Diagnostics for Model (4), including tests for pre-trend, placebo effect, and carryover effect, are in Web Appendix D.

Table 3. Estimation of Causal Impact of CPF Badge on Price

DV: Log (price+1)	**Model (3)**		**Model (4)**	
	ATT	p	ATT	p
Treated observations equally weighted	0.0580	0.0000	0.0535	0.0000
Treated units equally weighted	0.0492	0.0000	0.0447	0.0000
	$\hat{\beta}$	p	$\hat{\beta}$	p
Log (sales rank)	0.0128	0.0000	0.0165	0.0000
Log (rate # + 1)	−0.0074	0.0002	−0.0007	0.7504
Mean review positivity			−0.0932	0.0702
Mean rating score			0.0048	0.2956
Log (sustainability topic + 1)			−0.0210	0.0000
Log (packaging topic + 1)			−0.0276	0.0000

Description valence		0.0366	0.0081
Log (description length + 1)		0.0144	0.0873
Description richness		0.0878	0.0445
Description readability		0.0004	0.0213
Colorfulness		0.0001	0.4947
Brightness		−0.0142	0.6012
Symmetry		0.0000	0.2045
Face #		0.0456	0.0115
Image technical score		−0.0072	0.3816
Image aesthetic score		−0.0043	0.7871
RMSE	0.2043	0.2043	
Obs. #	1,274,958	1,274,958	

Notes. For identification purposes, units whose number of untreated periods was less than 5 were dropped automatically.



(a) ATT in Model (3) (b) ATT in Model (4)

Figure 7. ATT of Having the CPF Badge on Price from Model (3) [a] and Model (4) [b]

Notes. The gray bar denotes the number of units at the *t* period after treatment.

CPF Badge Adoption Leads to Decreased Market Concentration

We examined how adopting the CPF badge on one or more of a brand's products affected market concentration on Amazon. First, we analyzed the impact of badge adoption on demand for large and small brands separately. To define brand size, we used sales rank quantiles from March 1, 2023 to March 15, 2023 and divided brands within each category into three sizes—small, medium, and large—based on the 0.25 and 0.75 quantiles. We then replicated Model (2) from Table 2 for each brand size. The results showed that adopting the CPF badge helped small brands become more competitive, with a more pronounced effect than on medium-sized brands

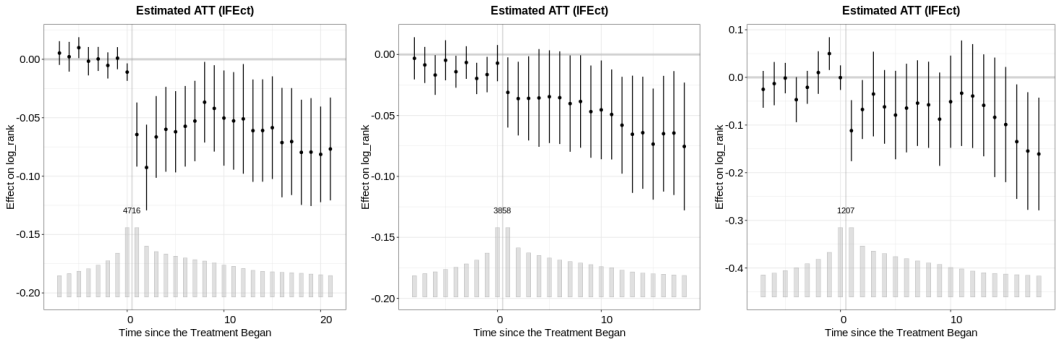
(Table 4). However, the CPF badge might increase market concentration, as the demand effect was strongest for large brands (note the different scales on the y-axis in Figure 8).

Table 4. Estimation of Causal Impact of CPF Badge on Demand by Brand Size (DV: Log Sales Rank, lower = higher demand)

	Small Brand		Medium Brand		Large Brand	
DV: Log (sales rank)	**ATT / $\hat{\beta}$**	**p**	**ATT / $\hat{\beta}$**	**p**	**ATT / $\hat{\beta}$**	**p**
Treated observations equally weighted	−0.0893	0.0000	−0.0433	0.0283	−0.1640	0.0011
Treated units equally weighted	−0.0786	0.0000	−0.0550	0.0008	−0.1557	0.0003
Log (price + 1)	0.0414	0.0909	0.1555	0.0000	0.1482	0.0000
Log (rate # + 1)	−0.3923	0.0000	−0.5456	0.0000	−0.5470	0.0000
Mean review positivity	0.2829	0.5875	0.4583	0.2236	0.1905	0.7737
Mean rating score	0.0454	0.4116	0.0552	0.2006	0.0409	0.4363
Log (sustainability topic + 1)	0.2895	0.0032	0.2966	0.0000	0.2784	0.0000
Log (packaging topic + 1)	−0.0106	0.9034	0.0546	0.2683	−0.0341	0.3497
Description compound	0.0533	0.8589	−0.1325	0.1695	−0.2217	0.0627
Log (description length + 1)	−0.2838	0.0715	−0.0924	0.0463	−0.0535	0.3809
Description richness	0.1575	0.7333	−0.6014	0.0677	−0.4232	0.0772
Description readability	−0.0041	0.2056	−0.0012	0.2300	−0.0022	0.0185
Colorfulness	−0.0004	0.8489	0.0021	0.1128	−0.0025	0.0529
Brightness	0.5150	0.4772	0.0317	0.9235	0.1243	0.7301
Symmetry	0.0000	0.3633	0.0000	0.8757	0.0000	0.3450
Face #	−0.4627	0.1708	−0.1926	0.1614	−0.2833	0.1202
Image technical score	0.2767	0.1383	0.0367	0.6440	0.1588	0.2021
Image aesthetic score	−0.1563	0.6124	−0.0317	0.7791	0.1441	0.1921
RMSE	0.2841		0.3269		0.4926	

Observations #	313,102	609,116	243,682

Notes. For identification purposes, units whose number of untreated periods was less than 5 were dropped automatically.



(a) ATT for Small Brands (b) ATT for Medium Brands (c) ATT for Large Brands

Figure 8. ATT of CPF Badge on Demand for Small [a], Medium [b], and Large Brands [c]

Next, we examined the relationship between CPF badge penetration (i.e., the percentage of products that had adopted the badge) and market concentration. We used the HHI (defined in the Equilibrium Outcome section) to measure market concentration. Given the potential exit and entry of small brands, we focused only on medium- and large-sized brands, which were relatively stable, and assumed that the market share held by small brands could be ignored. Additionally, we used absolute sales rank instead of relative sales rank among these medium- and large-sized brands.

To calculate the HHI, we used Equation (39). To calculate market share, we needed sales volume data. However, we had only sales rank data, and since Amazon does not disclose how it calculates sales ranks based on sales volume, we relied on the empirical transformation formula (He and Hollenbeck 2020) estimated with large datasets from Amazon for each category (not including the category of grocery and gourmet food). Finally, in our data for analysis, there were 75% medium- and large-sized brands from March 15, 2023 to September 15, 2023.

We calculated the daily HHI for the market of 20 subcategories and explored how the proportion of CPF-badged products (i.e., the badging threshold in the game theory model) in a subcategory influenced the HHI with a fixed effect model:

$$HHI_{st} = b_{s0} + b_1 CPFProportion_{st} + b_2 MeanPrice_{st} + \boldsymbol{b}_3\overrightarrow{Ratings_{st}} + \boldsymbol{b}_4\overrightarrow{Reviews_{st}} + \boldsymbol{b}_5\overrightarrow{Descriptions_{st}} + \boldsymbol{b}_6\overrightarrow{Images_{st}} + \tau_t + \varepsilon_{st}, \quad (45)$$

where $CPFProportion_{st}$ denotes the proportion of CPF-badged products in subcategory s at time t; and $MeanPrice_{st}$ was calculated by the mean of $log(price + 1)$ for products in category s at time t. Furthermore, $Ratings_{st}$ includes *log (rate count + 1)*, mean rating score; $Reviews_{st}$ includes the mean review positivity score and occurrences of sustainability-, packaging-, and return-related topic words; $Descriptions_{st}$ includes the mean compound valence score, length, richness, and readability of all product descriptions; and $Images_{st}$ includes the mean colorfulness, brightness, symmetry, average face count, technical quality score, and aesthetics quality scores of product images. We also controlled for subcategory fixed effect, b_{s0}, and time fixed effect, τ_t.

The estimation results for this fixed effect model showed a clear significant negative effect of the CPF proportion on the HHI, indicating that as more products adopted the badge, the market became less concentrated (Table 5). Even though we focused on medium- and large-sized brands, assuming small brands' market share was negligible, the conclusion holds: CPF badge adoption led to decreased market concentration, even if new small brands entered the market.

Table 5. Effect of CPF Badge Proportion on Herfindahl–Hirschman index (HHI), higher value = greater market concentration)

	Estimate	SD	p
CPF-badged product proportion	−324.3098	102.9078	0.0016
Log (price + 1)	383.7037	255.4267	0.1331
Log (rate # + 1)	98.6776	139.6599	0.4799
Mean review positivity	−50,201.1990	5,382.3200	0.0000
Mean rating score	3,908.9723	350.3440	0.0000
Log (sustainability topic + 1)	1,752.4851	346.7690	0.0000

Log (packaging topic + 1)	190.2718	350.2501	0.5870
Description valence	882.6418	466.1405	0.0584
Log (description length + 1)	−78.1181	149.7589	0.6020
Description richness	1,332.6806	761.2951	0.0801
Description readability	−20.7897	2.4830	0.0000
Colorfulness	16.2890	1.2295	0.0000
Brightness	2,593.9950	330.5784	0.0000
Symmetry	0.0040	0.0024	0.0976
Face #	−1,280.8494	92.8675	0.0000
Image technical score	135.3650	122.6360	0.2698
Image aesthetic score	−1,242.3769	119.7294	0.0000
(Intercept)	−410.7808	1,951.6722	0.8333
Observations	3,204		
R^2	0.8592		

Notes. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

DISCUSSION

We examined how a single, third-party green badge affected demand, pricing, and market concentration in online marketplaces by pairing a game model with causal evidence from real data. We first developed a three-stage game theory model in which the marketplace set a badge threshold at one of three levels (i.e., 0, 0.5, or 1) for green products. Furthermore, the sellers set prices to maximize profits, and consumers make purchase decisions, leading to profits for both the marketplace and the sellers. Our model derives a subgame perfect equilibrium for each badging threshold, identifying when badge adoption leads to increased demand, price, and market competition. These conditions held when eco-conscious consumers constituted the majority. The optimal badge proportion converged toward 50%, maximizing profits, particularly when the baseline utility of non-sustainability-related features in badged products was at least twice that of unbadged products, the product differentiation between sellers was sufficiently high, and green consumers significantly outnumbered non-green consumers. Empirical data support this hypothesis, showing that a badge threshold close to 50% correlates with increased demand, higher prices, and reduced market concentration.

Next, we analyzed daily data from March 1, 2023 to September 15, 2023 for 6,606 products across 8 categories and 20 subcategories. Using computer vision and deep learning, we extracted image features such as colorfulness, brightness, and image quality and used natural language processing techniques to analyze text features including valence, richness, and readability. We then employed multimodal machine learning models to create vector representations of images and descriptions, ensuring robust feature selection. Our data revealed that 35.39% of products had the CPF badge, supporting our hypothesis. To ensure robust causal inference, we used the IFEct estimator to address endogeneity issues and identify the CPF badge's impact on demand, price, and market concentration.

We found that the CPF badge significantly improved sales rank, indicating a positive effect on sales volume, although the impact was relatively short-lived, highlighting the importance of product imagery and descriptions. Additionally, we found that the CPF badge consistently raised prices, which is in line with previous research showing consumers' willingness to pay a premium for sustainable products. Interestingly, the CPF badge enhanced the market presence of small brands more than that of medium-sized brands; however, the badge may also increase the dominance of larger brands, ultimately reducing market concentration and fostering a more competitive marketplace. These insights are valuable for both researchers and practitioners interested in sustainable product marketing dynamics.

Our study is a pioneering investigation into the broad impacts of a unified eco-label—the CPF badge—on demand, pricing, and market concentration within the diverse landscape of an e-commerce platform. We reveal how badge adoption can simultaneously increase demand and pricing while reducing market concentration, illustrating a triple-benefit outcome supported by empirical data. Our contributions are threefold: (i) We provide empirical evidence that CPF

badge adoption increases sales on the demand side; (ii) we demonstrate that on the supply side, it raises prices and reduces market concentration; and (iii) we are among the first to integrate empirical causal inference with game theory modeling. We developed a game theory model to explain the conditions under which these outcomes occur. Our key contribution lies in the empirical findings, particularly the novel evidence of the supply-side effects on pricing and market competition, with the theoretical model clarifying the mechanisms behind these results.

The combination of theory and empirical evidence in our study yields valuable insights. The theoretical model showed that under appropriate conditions—such as a strong presence of eco-conscious consumers, high baseline utility, and sufficient product differentiation—the optimal badging threshold was around 50%. This optimal level balanced the benefits of badge adoption, avoiding the extremes of 0% (no badges) and 100% (all badges), both of which would be less effective. Our empirical analysis found that the observed badging proportion was about 34%. Although somewhat lower than the theoretical optimum, this 34% threshold is much closer to the ideal of 50% than to the extreme cases. This suggests that in practice, platforms tend to adopt a badge distribution that is near optimal, which validates the conditions outlined in our theoretical model. The convergence of the empirical findings with our model's predictions reinforces that a moderate level of badge adoption is sufficient to drive increased demand, higher prices, and enhanced market competition. Thus, integrating both approaches not only strengthens our causal inference but also provides guidance for platform managers on setting effective badging policies.

Our study provides actionable insights for platform operators, brand managers, and policymakers regarding their pricing and product strategies to maximize demand following the adoption of a green badge. Our findings are beneficial for e-commerce platforms such as Amazon, enriching their understanding of the CPF badge's positive impact on expanding the

customer base for sustainably marketed products. Consequently, our work sets a precedent for other e-commerce and third-party online platforms considering similar initiatives. We demonstrate a detailed framework for navigating sustainable marketing strategies, ensuring that the adoption of unified eco-labels is both effective in fostering competition and beneficial in enhancing consumer engagement with green products.

DECLARATIONS

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no funding to report.

REFERENCES

Aspara, J., X. Luo, and R. Dhar (2017), "Effect of Intelligence on Consumers' Responsiveness to a Pro-environmental Tax: Evidence from Large-scale Data on Car Acquisitions of Male Consumers," *Journal of Consumer Research*, 27(4), 448–55.

Bai, Jushan (2009), "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77(4), 1229–79.

Baker, M. J. and G. A. Churchill (1977), "The Impact of Physically Attractive Models on Advertising Evaluations," *Journal of Marketing Research*, 14(4), 538–55.

Banerjee, Subhabrata Bobby, Easwar S. Iyer, and Rajiv K. Kashyap (2003), "Corporate Environmentalism: Antecedents and Influence of Industry Type," *JM*, 67(2), 106–22.

Bollinger, Bryan, Randi Kronthal-Sacco, and Levin Zhu (2024), "Sustainable Product Profit Potential and Availability," Available at SSRN 4546321.

Borin, N., D. C. Cerf, and R. Krishnan (2011), "Consumer Effects of Environmental Impact in Product Labeling," *Journal of Consumer Marketing*, 28(1), 76–86.

Brecko, Kristina and Yewon Kim (2024), "Sustainability and Strategic Differentiation: Rising Preferences and Divergent Brand Strategies in Unregulated Consumer Markets," Kilts Center at Chicago Booth Marketing Data Center Paper (forthcoming).

Callaway, Brantly and Pedro H. C. Sant'Anna (2021), "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 225(2), 200–30.

Ceylan, G., K. Diehl, and D. Proserpio (2023), "Words Meet Photos: When and Why Photos Increase Review Helpfulness," *Journal of Marketing Research*.

Chen, Y. and C. Chang (2013), "Towards Green Trust," *Management Decision*, 51(1), 63–82.

Delmas, M. A. and V. C. Burbano (2011), "The Drivers of Greenwashing," *California Management Review*, 54(1), 64–87. https://doi.org/10.1525/cmr.2011.54.1.64

Fairchild, Mark D. (2013), *Color Appearance Models*, Wiley.

Gobillon, Laurent and Thierry Magnac (2016), "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," *Review of Economics and Statistics*, 98(3), 535–51.

Gorn, Gerald J., Amitava Chattopadhyay, Tracey Yi, and Darren W Dahl (1997), "Effects of Color as an Executional Cue in Advertising: They're in the Shade," *Man. Sci.*, 43(10), 1387–400.

He, S., B. Hollenbeck and D. Proserpio (2022), "The Market for Fake Reviews," *Mar. Sci.*, 41(5), 896–921.

Johnstone, Micael-Lee and Lay Peng Tan (2015), "Exploring the Gap Between Consumers' Green Rhetoric and Purchasing Behaviour," *Journal of Business Ethics*, 132(2), 311–28.

Kelly, W. A. (1981), "A Generalized Interpretation of the Herfindahl Index," *Sou. Econ. J.*, 48(1), 50.

Kreitler, Hans and Shulamith Kreitler (1972), "The Model of Cognitive Orientation: Towards a Theory of Human Behavior," *British Journal of Psychology*, 63(1), 9–30.

Liu, Licheng, Ye Wang, and Yiqing Xu (2022), "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data," *Ame. Jour. of Pol. Sci.*.

Liu, X., D. Lee, and K. Srinivasan (2019), "Large-Scale Cross-Category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning," *JMR*, 56(6), 918–43.

Luchs, M. G., R. W. Naylor, J. R Irwin, and R. Raghunathan (2010), "The Sustainability Liability: Potential Negative Effects of Ethicality on Product Preference," *JM*, 74(5), 18–31.

Luchs, Michael G. and Todd A. Mooradian (2012), "Sex, Personality, and Sustainable Consumer Behaviour: Elucidating the Gender Effect," *Journal of Consumer Policy*, 35(1), 127–44.

Patrick, M., N. Kangun, and W. B. Locander (1978), "Environmentally Concerned Consumers–Racial Variations," *Journal of Marketing*, 42(4), 61–66.

Mai, Long, Hailin Jin, and Feng Liu (2016), "Composition-Preserving Deep Photo Aesthetics Assessment," in *2016 CVPR*, IEEE: 497–506.

Mazar, N. and C. Zhong (2010), "Do Green Products Make Us Better People?" *Psy. Sci.*, 21(4), 494–98.

Menon, Ajay and Anil Menon (1997), "Enviropreneurial Marketing Strategy: The Emergence of Corporate Environmentalism as Market Strategy," *Journal of Marketing*, 61(1), 51.

Murray, N., L. Marchesotti, and F. Perronnin (2012), "AVA: A large-scale database for aesthetic visual analysis," in *2012 CVPR*, IEEE: 2408–15.

Newman, George E., Margarita Gorlin, and Ravi Dhar (2014), "When Going Green Backfires: How Firm Intentions Shape the Evaluation of Socially Beneficial Product Enhancements," *Journal of Consumer Research*, 41(3), 823–39.

Olsen, Mitchell C., Rebecca J. Slotegraaf, and Sandeep R. Chandukala (2014), "Green Claims and Message Frames: How Green New Products Change Brand Attitude," *JM*, 78 (5), 119–37.

Parguel, B., F. Benoît-Moreau, and F. Larceneux (2011), "How Sustainability Ratings Might Deter 'Greenwashing': A Closer Look at Ethical Corporate Communication," *Journal of Business Ethics*, 102(1), 15–28.

Park, Sungsik, Man Xie, and Jinhong Xie (2023), "Frontiers: Framing Price Increase as Discount: A New Manipulation of Reference Price," *Marketing Science*, 42(1), 37–47.

Peattie, K. and S. Peattie (2009), "Social Marketing: A Pathway to Consumption Reduction?" *Journal of Business Research*, 62(2), 260–68.

Pieters, Rik, Michel Wedel, and Rajeev Batra (2010), "The Stopping Power of Advertising: Measures and Effects of Visual Complexity," *Journal of Marketing*, 74(5), 48–60.

Radford, Alec, Jong Wook Kim, Chris Hallacy, et al. (2021), "Learning Transferable Visual Models From Natural Language Supervision."

Reczek, R. W., J. R Irwin, D. M Zane, and K. R Ehrich (2018), "That's Not How I Remember It: Willfully Ignorant Memory for Ethical Product Attribute Information," *JCR*, 45(1), 185–207.

Reimann, Martin, Judith Zaichkowsky, Carolin Neuhaus, et al. (2010), "Aesthetic Package Design: A Behavioral, Neural, and Psychological Investigation," *JCP*, 20(4), 431–41.

Sen, Sankar and C. B. Bhattacharya (2001), "Does Doing Good Always Lead to Doing Better? Consumer Reactions to Corporate Social Responsibility," *Journal of Marketing Research*, 38(2), 225–43.

Singh, Nirvikar and Xavier Vives (1984), "Price and Quantity Competition in a Differentiated Duopoly," *RAND Journal of Economics*, 15(4), 546–54.

Sun, Liyang and Sarah Abraham (2021), "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics, 225*(2), 175–99.

Sussman, Abigail B., and Rourke L. O'Brien (2016), "Knowing When to Spend: Unintended Financial Consequences of Earmarking to Encourage Savings," *JMR*, 53(5), 790–803.

Talebi, Hossein and P. Milanfar (2018), "NIMA: Neural Image Assessment," *IEEE Transactions on Image Processing*, 27, 3998–4011.

Townsend, Claudia, and Suzanne B. Shu (2010), "When and How Aesthetics Influences Financial Decisions," *Journal of Consumer Psychology*, 20(4), 452–58.

Tully, Stephanie M., and Russell S. Winer (2014), "The Role of the Beneficiary in Willingness to Pay for Socially Responsible Products: A Meta-analysis," *Journal of Retailing*, 90(2), 255–74.

White, K., J. Hardisty, and R. Habib (2019), "The Elusive Green Consumer," *Harvard Business Review*.

Xu, Yiqing (2017), "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models," *Political Analysis*, 25(1), 57–76.

Zhang, S., D. Lee, P. V. Singh, and K. Srinivasan (2022), "What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features," *Management Science*, 68(8), 5644–66.

Zhou, B. and T. Zou (2023), “Competing for Recommendations: The Strategic Impact of Personalized Product Recommendations in Online Marketplaces,” *Marketing Science*, 42(2), 360–76.

Zhu, Levin Liewen (2024), “Essays on Sustainable Product Consumption.” PhD dissertation, Duke University.

CHAPTER V: CONCLUSIONS

This dissertation has shown that first impressions, delivered through faces, the visual distinctiveness of images, and at-a-glance informational badges, systematically move attention, choice, and price, and that repeated at scale they reshape reviews, entry, and market concentration. The central claim is that these cues are not mere creative flourishes: they are measurable, causally meaningful, and governable design variables. Across the three studies, I developed interpretable computer-vision measures (celebrity visual potential and visual uniqueness), paired them with behavioral validation and large-scale causal designs, and linked micro-responses to platform-level policy, particularly around badging.

The managerial implications follow directly. For faces, the CVP metric offers a grounded way to think about "who appears first." CVP captures charisma-related inferences beyond attractiveness or typicality and predicts outcomes in social and professional contexts. It is therefore useful, but it must be used with care. Organizations should obtain explicit consent for facial data collection and disclose intended uses; compliance and auditability are prerequisites to any application. In practice, CVP should be treated as one input among several, its external validations are positive but explain modest variance, so it complements rather than substitutes for skills, experience, and fit. Deployed responsibly, CVP can inform creator/endorser selection, serve as a covariate in field experiments to improve statistical power, guide the design of AI-generated spokespersons or educators, and help HR diagnose whether hiring relies excessively on facial signals. Because we find no relationship between CVP and protected attributes such as gender or race in our data, the measure does not inherently encode those biases, but it should never be a sole screening criterion and should be periodically audited for subgroup performance.

For images of things, the visual-uniqueness work turns vague advice to “stand out” into an actionable rule: distinctiveness pays until it doesn’t. Demand follows an inverted-U in uniqueness, with stronger gains when a listing also signals responsiveness and quality. That shape makes optimization concrete. Platforms can surface a uniqueness score in creator tools, use it as a feature in ranking and recommendation for users who seek novelty, and nudge providers toward the middle of the curve rather than its extremes. On the provider side, the tooling can be practical and light-weight: a heat-map overlay highlights which elements raise or lower uniqueness; small edits to room features or camera framing can move a listing toward the sweet spot; even curating the image set matters, dropping the least-unique image raises a property’s average uniqueness meaningfully. Consistency across a listing’s images helps, too: wild swings in style confuse viewers and dilute the first impression. Although the model was trained on Airbnb images from one market, the method, psychology-guided, unsupervised representation learning with human validation, readily ports to other domains once retrained on local data.

For informational cues, the Amazon badge study demonstrates that compressing complex sustainability information into a single, credible signal increases demand and prices and, under realistic consumer mixes, reduces concentration by helping smaller brands compete for attention. The policy lever here is coverage. Our model predicts that moderate coverage, around one-half of products eligible in a category, often maximizes platform profits under common conditions (strong presence of eco-conscious consumers, sufficient product differentiation, and high baseline utility of badged items). Empirically we observe coverage in the mid-30s, closer to that optimum than to the extremes of zero or universal badging, and causal estimates show sales and price gains alongside lower concentration. In practice, platform managers should set coverage

deliberately rather than letting it drift, articulate whether the objective is revenue, consumer welfare, or seller surplus, and monitor distributional impacts across brand sizes. Because the sales lift can be short-lived absent strong creative, managers should pair badging with guidance on imagery and descriptions. the informational cue and the visual cue are complements at first glance.

These implications add up to a simple operating playbook. Measure what matters at a glance with interpretable tools and human validation; design the first frame—who appears, how it looks, what it says—toward clearly stated objectives; deploy with experiments and instrumentation that track not just clicks and conversions but also prices, reviews, and concentration; and govern with audits for fairness, privacy, and credibility. The ethical obligations are non-negotiable: obtain consent for facial data; be transparent about labeling criteria; publish documentation on measurement and governance; and test for subgroup performance and potential harms.

The work is not without limits. External validity requires retraining and re-validation as contexts change; observational designs, however careful, cannot eliminate all confounding; and most analyses are partial-equilibrium snapshots that precede strategic adaptation by creators and sellers. These limits motivate the next steps: studying cross-cue interactions (faces × image style × badges), embedding experiments in allocation systems, extending to longer horizons to capture feedback via reviews and content evolution, and porting the methods across cultures and categories. As generative tools lower the cost of visual and textual optimization, research and practice should co-develop guardrails that preserve authenticity and trust.

In closing, the dissertation reframes first impressions as a system of levers rather than a matter of taste. Faces (CVP), visual distinctiveness, and badges are small, glance-level signals that can be measured and tuned. When platforms and sellers design these signals deliberately,

and disclose, audit, and adjust them, they can create clearer information environments, fairer and more effective persuasion, and healthier competition. Designing first impressions is designing the market's front door.