Using the wisdom of the crowd to improve the condition assessment of residential real estate

Routledge
Taylor & Francis Group

Journal of Real Estate Research

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rjer20

Using the Wisdom of the Crowd to Improve the Condition Assessment of Residential Real Estate

David Koch, Michael J. Seiler & Simon Thaler

JOURNAL OF REAL ESTATE RESEARCH
https://doi.org/10.1080/08965803.2024.2352223

OPEN ACCESS Check for updates

Using the Wisdom of the Crowd to Improve the Condition Assessment of Residential Real Estate

David Koch[a], Michael J. Seiler[b] and Simon Thaler[c]

[a]Institute for Energy, Facility & Real Estate Management, University of Applied Sciences Kufstein, Kufstein, Austria; [b]Raymond A. Mason School of Business, College of William and Mary, Williamsburg, VA, USA; [c]Henley Business School, Department of Real Estate and Planning, University of Reading, Reading, UK

ABSTRACT

The "wisdom of the crowd" refers to the phenomenon where by the average estimate of a group is superior to the estimate of any individual within the group, even if that individual is an expert and the others are not. When applied to real estate, it demonstrates that a group of non–real estate professionals can arrive at a more accurate physical property categorization than a real estate expert. Importantly, this more accurate assessment of the property's physical condition can be achieved by merely viewing photos of the property. This appraisal trend makes the finding even more relevant to practical implementation.

KEYWORDS

Wisdom of the crowd; valuation; behavioral experiment; appraisal; subjectivity bias

JEL CLASSIFICATION

R32; G41

Introduction

The wisdom of the crowd (WoC) refers to the idea that a group of people generally has more knowledge and insight than an individual. Galton (1907) was the first to describe this effect when he found that the average of all estimates of the weight of an ox was closer to the actual weight than any single estimate by an individual. This effect gained popularity through Surowiecki (2004) and van Dolder and van den Assem (2018), who successfully applied this principle to improve economic forecasts, medical judgments, and meteorological predictions.

We apply the wisdom of the crowd to real estate to classify the physical condition of a property. We focus on the property's condition because it is a physical characteristic that can be visually assessed and because it is a crucial factor influencing market value (Miller et al., 2018). Unlike other objective variables, such as age or square footage, this assessment is based on the appraiser's individual qualitative and subjective judgment, which is obtained from a visual inspection of the property. During the COVID-19 pandemic, appraisers relied heavily on property photographs due to restrictions on physically entering homes. This practice has now been widely integrated into the appraisal

CONTACT Simon Thaler simonthaler@gmx.net University of Reading, Henley Business School, Department of Real Estate and Planning, Whiteknights, Reading RG6 6UD, UK.

process. As such, appraisers often rely on photos of the inside of the home to make value judgments.[1] Given that visual representations are known to bias individuals' decision-making (Winkielman & Gogolushko, 2018), we examine whether the WoC can improve home price estimates and thus mitigate a potentially evoked subjectivity bias in individual condition assessments.[2]

People are generally viewed as having bounded rationality because they do not possess as much information as they desire and suffer from limited foresight regarding future events (Simon, 1955). In behavioral economics, the term "satisficing" is widely used to describe the willingness to accept a "good enough" solution, since cognitive limitations and time constraints make identifying a Pareto optimal solution unrealistic. However, when these imperfect assessments are aggregated from independently and identically distributed individuals[3] and diverse groups using a decentralized aggregation mechanism that fosters specialization, superior predictions are achieved as individual prediction errors are canceled out.[4]

Shanteau (1995) and Armstrong (1980) see no substantial advantage of expertise in individual predictions, as expert judgments are often inconsistent with those of other experts and even with themselves over time. In a real estate setting, appraisal smoothing bias is well established, and as such, it lends itself to an ideal testing ground to examine the wisdom of the crowd.

In this study, we conduct an experiment comparing physical condition assessments of 10,200 single-family homes made by non–real estate professionals with those made by a real estate valuation expert, using only photographs. This approach replicates the environment of "drive-by" or mass appraisals. The results demonstrate that the aggregation of property condition assessments (i.e., the wisdom of the crowd) stochastically dominates that of the expert appraiser after incorporating the views of 27 nonprofessionals in the Monte Carlo simulation. Essentially, the WoC effect, even when derived from individuals without real estate valuation experience, can mitigate the subjectivity bias inherent in the assessment of a single real estate professional.

Literature Review

The WoC phenomenon has been applied to various areas such as economic and political forecasts (Budescu & Chen, 2012; Mellers et al., 2014), public policies (Morgan, 2014), financial markets (Ray, 2006), nuclear safety (Cooke & Goossens, 2008), as well as to create actionable, spatially specific data for land use decisions (Brown, 2015). In WoC predictions, incorporating the opinions of individuals who are more knowledgeable about the subject improves the accuracy of the predictions (Murr, 2011). Moreover, a smaller group of experts can often make more accurate predictions than a larger crowd of uninformed individuals (Goldstein et al., 2014). However, it is important to acknowledge that biased individuals and the potential for information sharing or communication among participants can lead to erroneous results (Kao et al., 2018; Lorenz et al., 2011; Palley & Soll, 2019). Therefore, identifying underperforming individuals and excluding them from the crowd can be beneficial (Budescu & Chen, 2012).

The WoC is not limited to mere estimation tasks; it also plays a significant role in prediction markets. An example of this is the financial market's reaction to the space shuttle

Challenger disaster in 1986. The stock market quickly adjusted the share price of the implicated company within the first day following the accident (Maloney & Mulherin, 2003). This underscores several critical components necessary for the WoC to function optimally: a diversity of opinions that enriches the group's collective knowledge, obtained from individuals with private insights; independence in decision-making, highlighted by the lack of private communication immediately after the crash; decentralization, where individuals' specialized knowledge contributes to the overall understanding; and an effective aggregation mechanism, in this case, the stock market price, which translated private judgments into a collective conclusion. Remarkably, the stock market demonstrated superior predictive abilities than other mechanisms in determining the outcome of the ensuing investigation, delivering insights at a remarkably swift pace. This principle echoes the findings of the famous ox weight estimation experiment by Galton (1907), wherein the aggregate judgment of the crowd proved more accurate than that of any single individual.

In the realm of visual judgment and the WoC, research suggests that a simple majority voting mechanism is the most effective method of aggregation, yielding optimal results (Hastie & Kameda, 2005; Juni & Eckstein, 2017). Remarkably, even the combined decision of just two individuals can result in superior outcomes compared to singular judgments (Bahrami et al., 2010; Koriat, 2012). Given that the assessment of a home's condition involves evaluating multiple individual factors, it is important to recognize that the WoC effect extends to combinatorial problems. In such scenarios, various components come together to inform the final prediction (Yi et al., 2012). This suggests that leveraging the WoC in tasks requiring the integration of diverse pieces of information can significantly enhance decision-making accuracy.

Our research highlights the significant impact of the WoC effect across various scientific disciplines, while its application in the realm of real estate valuation is just beginning to be explored. We see a considerable opportunity in applying the WoC to assess the condition of real estate properties. Specifically, the condition variable not only enhances the accuracy of real estate pricing models but also is frequently overlooked in valuation processes (Miller et al., 2018). This oversight leads to renovation bias, which can distort property price indices by not fully accounting for the quality improvements from renovations, thereby potentially overstating overall price growth in time-series comparisons (Bogin & Doerner, 2019).[5] The significance of property quality and condition is underscored by Knight and Sirmans (1996), Wilhelmsson (2008), and Francke and van de Minne (2017), who note that neglected maintenance can significantly decrease a property's value over time (Francke & van de Minne, 2017), though it may not affect the likelihood of a sale (Billings, 2015). Additionally, the condition of a property not only affects its market value but also has a profound impact on the mental and physical health of its occupants (Palacios et al., 2021). Interestingly, despite the game-theoretical optimal decision being renovation rather than abandonment, new properties are often preferred over those requiring ongoing maintenance (Simmons-Mosley, 2003).

The inclusion of a condition variable in real estate valuation is significantly influenced by the party responsible for disclosure. Brokers' decisions on whether to disclose property condition defects are shaped by a variety of factors, including associated costs, prevailing market conditions, potential commissions, and legal considerations (Wiley &

Zumpano, 2008). The enactment of disclosure laws pertaining to property conditions serves to mitigate market risk premiums (Nanda & Ross, 2012). In addition, the architecture of a property (Ahlfeldt & Mastro, 2012; Buitelaar & Schilder, 2017; Lindenthal & Johnson, 2021) and its overall attractiveness (Johnson et al., 2020; Wang et al., 2019) significantly influence real estate prices. The role of property images is pivotal in valuation processes (Kostic & Jevremovic, 2020), often leading to elevated sales prices (Benefield et al., 2011).

While existing literature extensively explores the impact of property condition on pricing, there is a notable paucity of discussion concerning the direct assessment or estimation of condition itself. Research tends to concentrate on biases within real estate behavior, such as the influence of previous contract prices on appraisers' valuation decisions (Eriksen et al., 2020). Although visual inspections are crucial for precise valuation and effective marketing, they are susceptible to bias. In that regard, Hebdzyński (2023) finds that aligning textual and visual quality signals through methods such as supervised machine learning algorithms can help bridge these gaps, with a notable consistency observed between the two in both sales and rental markets, suggesting a strategy to improve real estate market analyses.

Presenting a property neutrally, particularly in visual terms, poses a challenge. Various elements, including emotionally charged images or variations in lighting, can significantly affect decision-making processes (Lakens et al., 2013; Orquin et al., 2018; Winkielman & Gogolushko, 2018). These biases extend to the real estate market, where disparate presentations of an identical property result in fluctuating or biased valuations (Gillingham & Watten, 2024; Koch et al., 2021; Luchtenberg et al., 2019; Thaler & Koch, 2023). Moreover, a property's valuation is influenced by visual perceptions of its surroundings and neighborhood, encompassing aspects such as home upkeep, the dimensions of adjacent buildings, and the general state of the neighborhood (Ding et al., 2000; Leguizamon, 2010; Pavlov & Blazenko, 2005; Turnbull et al., 2006; Zabel, 1999). The proximity to mixed-use developments (Nakamura et al., 2018), waste facilities (Aydin & Smith, 2008; Ready, 2010), or green spaces, whether public or private gardens, also impacts property values via visual cues (Turner & Seo, 2021; Voicu & Been, 2008).

In summary, existing research underscores the promising application of the WoC in residential real estate valuation. Additionally, these studies highlight the critical importance of property condition in valuation processes and point out the significant distortions that visual factors can introduce into decision-making processes.

Methodology

Our methodological approach is fundamentally based on an economic experiment. Specifically, we engaged 102 nonprofessionals and one professional appraiser,[6] asking them to classify the physical condition of 100 single-family homes into four distinct condition classes. The participants were full-time and part-time students, 54 males, 47 females, and one individual identifying as nonbinary. Their ages ranged from 18 to 54 years, with an average age of 27. Of these students, 37 worked full-time, 40 were part-time employees, and 21 were unemployed. They all attended different courses in which they conducted the experiment. At the beginning of the lecture, they entered a computer laboratory where they conducted the experiment online on separate computers. The real estate professional was a court-certified appraiser with the highest level of certification and extensive experience in both industry and court settings, making him highly familiar with the market and its conditions. He conducted the experiment at a different time, separately from the students and independently of their results.

In the upcoming sections, we provide an overview of our experimental design and elaborate on the four classes of house condition as defined by official appraisal standards. We then analyze the outcomes of these classifications through a hedonic pricing model to assess their influence on property values. Our methodology is divided into three primary segments: (1) an overview of the experimental framework and the detailed criteria for the four condition classes; (2) an insight into the specific execution of the experiment and the incentives offered; and (3) a detailed description of our methodology for data analysis.

Structure of the Experiment

Our experimental method involves recruiting nonprofessionals to undertake a comprehensive series of tasks designed to gauge their perceptions regarding the condition of single-family homes. A thorough explanation of the experiment is outlined in Figure 1.

Task 1 (T1) – Demographic Questions

Participants begin the experiment by answering a series of demographic questions in Task 1. These questions cover a range of topics, including their field of study, employment status, gender, and age, with the objective of establishing a comprehensive profile of the diverse backgrounds represented in our participant pool. To ensure ease of response and clarity, the questionnaire is designed in a user-friendly manner, utilizing either a radio matrix or a checkbox format.

Task 2 (T2) – Visual Comparison of Condition Classes

In Task 2, participants engage with a dynamic visual exercise designed to deepen their understanding of the four property condition classes. Each class is represented by two images, accompanied by detailed descriptions that highlight the characteristics unique to each class, as defined by the official guidelines of the Association of Appraisers. Participants are asked to compare these images and articulate the perceived differences between the classes. This task is facilitated through the use of a radio matrix, allowing participants to rate the level of dissimilarity on a scale from 1 (no discernible difference)

Task	T1	T2	T3
Treatment A		/	A1 - A5, 6 - 95, B1 - B5
Treatment B		/	B1 - B5, 6 - 95, A1 - A5

Figure 1. Tabular display of conducted treatments and their consecutive tasks. Both treatments start by asking demographic questions ([[INLINE FIGURE]]) and questions regarding real estate employment and experience in Task 1 (T1). In Task 2, participants have to assess the difference among the four condition classes based on two sample images per class (T2). The next task (T3) depicts 100 pictures beginning with five treatment photos that show a property in different states (two black and white, zoomed, mirrored, and duplicate). The opposite effect is shown again at the end of this task.

to 7 (a significant difference). An example of this exercise is detailed in Appendix Figure A1.

Main Task (T3) – Property Assessment by Condition Classes

Task 3 represents the key component of the experiment, where participants are presented with a series of 100 screens. On each screen, they are tasked with assessing the condition of a property based on a single front-view image. To aid in their evaluation, a detailed table describing the four condition classes is provided, guiding participants to focus solely on the property's condition. The classes are organized from highest quality (Class 1) to lowest quality (Class 4). Considering the potential for diminished attention over an extended period, and in an effort to maintain a feasible duration, the experiment is designed with 100 images. Participants are expected to spend an average of 10 to 15 seconds reviewing each picture, resulting in a total completion time of around 20 minutes.

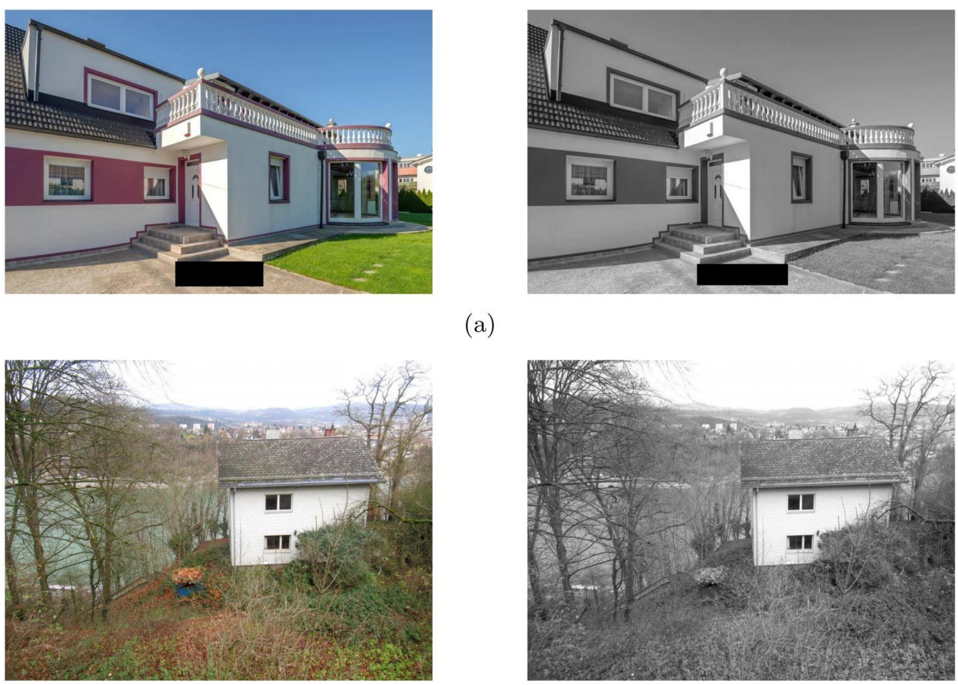
The classification system used in the experiment is aligned with various building components, such as construction quality, roofing, facades, and windows and exterior doors, condensed into the four categories. Class 1 denotes properties with high-quality construction materials, recent construction years, and sophisticated design elements, representing the highest standard. Class 2 corresponds to properties with medium-quality materials, showcasing solid construction without necessarily being new, and with minimal depreciation. Class 3 describes properties with lower-quality materials, older construction, or those in need of minor renovation. Finally, Class 4 applies to properties that either are on the brink of demolition or require extensive renovations, with the renovation costs outweighing the building's value, thus highlighting the land's value as the primary consideration. Participants are tasked with categorizing each home into one of these four classes, using a radio matrix located beneath the photo of the property. For a depiction of this classification process, refer to the illustrative screenshot in Appendix Figure A2.

The employed dataset comprises a collection of 92 single-family houses, all sourced from an online real estate brokerage platform featuring properties for sale in Austria. The selection encompasses a wide variety of house types—including single-family homes, townhouses, and luxury villas—spanning Austria's geographic area, from major cities such as Vienna, Graz, and Linz to more rural locations. This diversity ensures a comprehensive representation of different construction dates, sizes, and regions within Austria, offering a nuanced understanding of the country's real estate landscape. The data collection period extends from January 2020 to August 2020. Notably, homes constructed after 2018 were excluded to eliminate any potential bias toward newly built homes typically sold on the primary market.

In addition to images, we selected key housing characteristics such as municipality, lot size, building size, year of construction, and pricing information. In our analysis, the front image, building characteristics, and aggregated land prices are utilized as independent variables in a hedonic regression model. The properties chosen for the experiment were specifically selected to minimize the visibility of their surroundings, with the exception of the pool treatment image, in an effort to reduce the environmental impact on valuation and mitigate potential biases due to subjective judgment. Our focus is primarily on the structural quality/value rather than the view's value, as incorporating the entire property value would necessitate considering the property's surroundings.

Participants were randomly divided into two groups, each exposed to a unique treatment. This involved varying the treatment and control images placed at the beginning and end of the task. The treatment images featured specific modifications such as zooming, grayscale conversion, altered color intensity (high intensity in pink), redundancy, and image mirroring to investigate potential biases in the participants' classification judgments. Additionally, the experiment incorporated duplicate control images to assess the consistency of participants' responses.

Figures 2 and 3 showcase the treatment images used in the experiment, including two color photographs and one zoomed-in image that notably omits much of the property's surroundings, particularly the prominent pool in the garden. These images aim to assess the impact of subtle visual modifications on participants' classification judgments. Recognizing that individuals' decisions can be influenced by visual information biases, our experiment seeks to determine whether such minor image manipulations can significantly affect classification outcomes. Of the 92 homes featured in the study, eight appeared twice to further explore biases; four of these duplicates were altered (two in black and white, one zoomed, and one mirrored) to delve deeper into the nuances of our primary research question.



(a)

(b)

Figure 2. Color treatment.
The images in (a) show a Class 2 house painted in a recognizable color (treatment color, left) and depicted in black and white (treatment black and white, right). The images in (b) show a Class 3 house which is painted in a recognizable color (treatment color, left) and depicted in black and white (treatment black and white, right).

Figure 3. Zoom treatment.
The images show a Class 2 house that is not zoomed in and a pool (treatment unzoomed, left) and a house that is zoomed in and no pool (treatment zoomed, right).

This experiment is designed to examine the complex relationship between visual stimuli, participant decision-making processes, and potential biases in evaluating property conditions. Through a detailed methodological approach, we seek to provide meaningful contributions to the understanding of visual perception's role in real estate assessment and appraisal.

Implementation and Incentivization

Recall that the WoC effect can be harnessed using individuals who are not professionals in the relevant field. Surowiecki (2004) argues that decentralization involves the specialization of group members, where some individuals may possess varying degrees of specialized or general knowledge or experience. This diversity within the crowd tends to reduce the group's overall estimation error (Page, 2008). To illustrate the potency of the WoC, we assembled a varied group of students, some of them unrelated professionally to real estate (N = 48) and other part-time students (N = 54) who have accrued professional and educational experience in the real estate sector. Despite criticism about using students in studies in which unfamiliarity with the subject matter and the need for life experiences and specific skills are concerns, our choice aims to explore the WoC's effectiveness across a broad demographic spectrum. Our subjects range in age from 18 to 54, with diverse levels of professional training and life experience. This varied composition allows us to examine the resilience and applicability of collective judgment in contexts that do not strictly depend on specialized knowledge. Hence, it showcases the WoC's utility in real estate evaluation by practically anyone. In line with this approach, we gathered property condition classifications from a cohort of 102 students.

To encourage focus and optimal effort on the tasks presented, subjects receive a monetary incentive. At the start of the experiment, the experimenter states that participants can earn up to 25 EUR. They are guaranteed 5 EUR for completing all questions in the experiment. After the tasks are explained, it is clarified that compensation is performance based, utilizing a coordination game–like payment scheme. In this scheme, payment depends not only on an individual's assessment of the correct answer but also on consensus within the group. This approach is inspired by the Keynesian Beauty Contest (Keynes, 1936), with coordination games detailed further in Schelling (1980). The final payment is determined through an aggregation mechanism, where the decision is influenced by the choices of all other participants, similar to how economic free markets aim to allocate resources to their most effective uses (Chamberlin, 1948; Smith, 1962). This setup reflects the dynamics of the real-world property market, where opinions on the condition of a house are formed independently by every participant.

The incentive scheme operates as follows: to encourage participants to provide their most accurate assessment of a property's condition, we rank them based on how closely their answer aligns with the group's mode response. If a participant's answer matches the mode of the group (with an average group size of 20 subjects), they earn a point. At the experiment's conclusion, we calculate each participant's score by comparing their individual answers to the group's mode response. Subsequently, we rank the answers, with the highest scores indicating the top performers. The top quintile receives an additional 20 EUR on top of the base participation fee of 5 EUR. Subsequent quintiles are awarded 15 EUR, 10 EUR, and 5 EUR, respectively, with the lowest quintile receiving only the participation fee. Payments are made discreetly in cash after the experiment concludes. An experimenter and the participant finalize the payment process privately, including the distribution of payment receipts, in a separate room.

Data Analysis

After the conclusion of the experiment, we aggregate all 10,200 property condition assessments to determine the overall mode condition rating for each of the single-family homes. This process yields a WoC condition evaluation for all 92 homes in our dataset, which we then use as an independent variable in our hedonic pricing model. Initially, we develop a baseline hedonic pricing model that includes all available property features, such as land value (expressed in thousands), square footage, and construction age. A second model is then constructed to incorporate the property condition as well.[7] Due to the availability of data on the age variable, this model is based on 74 observations. By comparing the two models, we evaluate the impact of adding a property condition variable on the model's accuracy, comparing both models to individual estimates and the collective judgment of the WoC.

Our baseline Model 1 (a) contains the following variables: land value, square footage of the property, and a second-degree polynomial function for the year of construction.[8] For the dependent variable, we utilize the logarithm of the price. In our regression analysis, we include only homes that are unique to our dataset, specifically excluding those featured in the treatment images and ensuring they have a documented year of construction.

In the treatment models, we include the property condition assessment in three different designs: Model 1 (b) incorporates the condition assessment of the real estate professional, Model 1 (c) includes the WoC assessment (the mode of the 102 subject assessments), and Model 1 (d) uses the individual assessment of each participant. The average coefficient of determination of the individual assessment of subjects is then compared to the WoC coefficient and the coefficient in the professional appraiser model. We also examine the general effect of the inclusion of a condition variable based on the comparison of the baseline model and the three models including a property condition variable.

To maintain the integrity of our analysis, three duplicate photos, two in black and white, one mirrored, and one zoomed, are omitted from the primary hedonic pricing model. These images are designated as treatment images and, if included, could lead to inconsistencies in the analysis.

$$\begin{aligned}
&(a)\ \log(Price) = f(poly(year,2), land_value, square_footage)\\
&(b)\ \log(Price) = f(poly(year,2), land_value, square_footage\\
&\qquad + Condition_Expert)\\
&(c)\ \log(Price) = f(poly(year,2), land_value, square_footage\\
&\qquad + Condition_WoC)\\
&(d)\ \log(Price) = f(poly(year,2), land_value, square_footage\\
&\qquad + Condition_Individual)
\end{aligned} \tag{1}$$

After estimating the regression models, we analyze the individual responses from subjects. This involves analyzing the adjusted R-squared for each subject based on their independent property condition assessments. Our aim includes to gathering insights on the minimal crowd size that produces results superior to those of the professional appraiser. To achieve this, we conduct a Monte Carlo simulation with 10,000 iterations, randomly selecting from the total pool of subjects without replacement.[9] We systematically increase the number of subjects included in the WoC regression (from 1 to 102 subjects). For each increment, we run 10,000 simulations to calculate a collective condition assessment, which is then used as the condition variable for the individual buildings in our regression model.

The primary focus is to observe how the adjusted R-squared fluctuates with an increasing number of subjects, aiming to identify the moment the collective accuracy exceeds that of the professional appraiser. In essence, this methodological approach helps us identify the crowd size required to achieve the WoC effect, specifically determining the quantity of nonprofessional individuals needed to outperform the expert's precision.

Last, to investigate whether deviations in the mode assessments of subjects—the WoC assessment—improve the hedonic model, we compare Model 1 (b) with a model based on Equation (2). This model introduces a factorial variable for deviations between the expert assessment and the WoC assessment into the hedonic model based on the expert's assessment:

$$\begin{aligned}
\log(Price) = f(&poly(year,2), land_value, square_footage + Condition_Expert\\
&+(Condition_Expert - Condition_WoC))
\end{aligned} \tag{2}$$

The two models reveal whether incorporating additional information derived from the WoC assessment significantly enhances the model's quality. Since the models are nested, we use an F-test to determine the impact of including the deviation variable.

Results

The difference between the WoC and professional assessments forms the foundation of the hedonic model. When analyzing the dataset, which includes the year of construction, we observe an accuracy of 68.9%. The confusion matrix detailed in Table 1 shows that 20 pictures are rated in a higher class by the WoC than by the expert's assessment. In

Table 1. Confusion matrix of the construction year restricted dataset.

WoC Classification		Expert Classification Class 1	Class 2	Class 3	Class 4	Total
	Class 1	7	5	0	0	12
	Class 2	1	22	10	0	33
	Class 3	0	1	17	5	23
	Class 4	0	0	1	5	6
	Total	8	28	28	10	

The number of matching class assignments between expert and WoC assessments are on the main diagonal. The lower class assignments of the WoC are below the main diagonal, and higher class assignments are above the diagonal.

contrast, three properties are rated in a lower class by the WoC compared to the expert's assessment, underscoring the assessment variation between the WoC aggregation mechanism and a single expert. In examining the discrepancies in the professional and nonprofessional assessments across all individual photos, we note that 66% of the 100 images are classified in the same category. Out of the 34 images classified differently, nonprofessionals place 27 of these homes in a class one level higher, while six images are deemed one class lower. There is a single instance of a home with a two-class deviation.

The subjects' tendency to overestimate highlights a crucial aspect of the WoC mechanism: while it aggregates diverse perspectives, it is not free from potential systematic biases such as overoptimism and anchoring, which can cause valuations to skew upward. However, this discrepancy also opens the possibility of a biased condition estimation by the professional appraiser. The pattern of subjects' overestimation relative to the appraiser's assessments carries significant theoretical implications. It suggests that the WoC is capable of leveraging diverse inputs to produce accurate outcomes under specific conditions and hints at the possibility of individual expert biases leading to systematic errors in current appraisal practices. This observation is consistent with the literature on cognitive biases and decision-making among appraisers and could represent an initial step toward mitigating such biases.

The practical implications of our findings suggest that crowd overestimation could lead to inflated property prices, aligning more closely with market values but potentially distorting market dynamics and impacting affordability. Therefore, our research highlights the essential need for a carefully considered application of WoC principles in real estate valuation. While the collective intelligence of the crowd offers a promising avenue for democratizing and enhancing property assessments, it is crucial to acknowledge and mitigate its limitations. To leverage the benefits of the crowd effectively while preventing potential overestimation, the introduction of checks, such as expert oversight and algorithmic adjustments, is vital. Future research should evaluate additional measures for controlling and mitigating these effects in more depth, providing guidance on refining WoC applications in real estate and other fields.

The variability in the class assignments by the WoC, as shown in Table 2, plays a significant role. The data reveal that, on average, Classes 1 and 4 receive 66% and 76% of votes, respectively, indicating that images categorized as Class 4 are more distinctly recognized by participants. Moreover, Class 2 images are equally prone to being misclassified as Class 1 or Class 3, whereas Class 3 images are less frequently mistaken for Class 4

Table 2. Variation in the WoC class assignments.

	WoC Classification			
	Class 1	Class 2	Class 3	Class 4
% Votes Class 1	66.3	18.1	0.6	0.0
% Votes Class 2	31.7	63.0	18.0	1.3
% Votes Class 3	2.0	18.5	68.8	22.7
% Votes Class 4	0.0	0.4	12.5	76.0
Standard Deviation	14.8	9.4	9.9	16.8
N	12	33	23	6

The columns indicate the classes assigned by the WoC aggregation mechanism. The rows % Votes Class 1 to 4 show the mean votes per class in the respective WoC classes. The row Standard Deviation depicts the variation of the absolute votes in the WoC Classes 1 to 4.

properties. Notably, despite exhibiting the highest standard deviation in vote counts, Class 4 images are most consistently identified correctly. This pattern underscores the varying degrees of clarity with which different classes are perceived by the participants.

As demonstrated in Table 3, our investigation into the WoC effect on property valuations employs a straightforward hedonic model. By incorporating the mode property condition assessment (i.e., the WoC condition classes) alongside variables such as land value, square footage, and year of construction, our WoC hedonic regression model achieves an adjusted R-squared of 0.728. When we compare this figure with the adjusted R-squared values of all individual assessment models—which range between 0.564 and 0.721, with an average of 0.623—we observe that the WoC model outperforms all individual models. The distribution of all the adjusted R-squared values is shown in Figure 4. Interestingly, regardless of the subjects' level of expertise, any classification into any of the four condition classes positively impacts the model's adjusted R-squared, with even the lowest-performing assessment slightly improving the model's fit compared to the baseline regression, which has an adjusted R-squared of 0.555. Additionally, when comparing assessments made by real estate students to those made by students in other study programs, no significant difference in assessment precision relative to the expert's is observed, indicating that the WoC effect holds across varying levels of subject expertise.

In comparison, the hedonic model that utilizes the assessment by the real estate professional yields an adjusted R-squared of 0.701. This outcome suggests that the condition assessment based on all nonprofessional participants, or the WoC, enhances the model's performance. A comprehensive model that includes all 92 assessed properties from the entire available dataset (excluding the year of construction) is provided in Appendix Table A2. Upon analyzing the coefficients associated with the three dummy variables for condition Classes 2 through 4, we observe a pattern consistent with the findings of Miller et al. (2018) and Palacios et al. (2021). Specifically, the lower the condition classification, the greater the discount applied. Essentially, properties of lower quality are subject to more significant value deductions. When we compare the expert's assessment with that of the WoC, as illustrated in Table 3, it is evident that the discount for properties in Classes 2 and 4, expressed in percentage terms due to the semi-log model, is more pronounced under the WoC framework. Conversely, for Class 3 properties, the deduction in value is similar across both the expert and WoC assessments, highlighting the nuanced differences in valuation adjustments between expert and collective assessments.

For further interpretation of the condition class coefficients, consider the exponentiated coefficient of the expert Class 2 coefficient in Model (b) of Table 3, which is

Table 3. Baseline and property condition pricing models for the professional and WoC estimates.

	Dependent Variable:		
	log(PRICE)		
	(a)	(b)	(c)
LAND VALUE [1K]	0.002***	0.002***	0.002***
	(0.0003)	(0.0003)	(0.0003)
SQUARE FOOTAGE	0.004***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)
poly(YEAR, 2)1	2.098***	0.393	0.264
	(0.479)	(0.491)	(0.470)
poly(YEAR, 2)2	0.391	−0.387	−0.314
	(0.469)	(0.421)	(0.405)
EXP_CLASS 2		−0.421**	
		(0.166)	
EXP_CLASS 3		−0.927***	
		(0.186)	
EXP_CLASS 4		−1.090***	
		(0.223)	
WOC_CLASS 2			−0.546***
			(0.135)
WOC_CLASS 3			−0.911***
			(0.168)
WOC_CLASS 4			−1.402***
			(0.214)
Constant	11.969***	12.809***	12.835***
	(0.153)	(0.221)	(0.185)
Observations	74	74	74
R^2	0.579	0.730	0.754
Adjusted R^2	0.555	0.701	0.728
Residual Std. Error	0.464 (df = 69)	0.380 (df = 66)	0.363 (df = 66)
F statistic	23.757***	25.438***	28.865***

Note: *p <0.1; **p <0.05; ***p <0.01.

This exhibit shows results for the baseline regression as well as for the two regressions with the property condition assessment. The dependent variable in the baseline regression Model (a) is the log of the home price. The independent variables include *land value*, *square footage*, and the polynomial of the year of construction (*year*). In the professional Model (b), a dummy for the four condition classes is added (*exp_class*). In the WoC Model (c), the regression is extended by including a dummy for the mode subject's assessment (*WoC_class*). The reference class in both models is the highest classification (Class 1).

calculated as $\exp(-0.421) = 0.656$. This result indicates that moving from the reference Class 1 to Class 2 is associated with a decrease in property value by approximately 34.4%, calculated as $(\exp(\beta) - 1) * 100 = (0.656 - 1) * 100$. For Classes 3 and 4 this would result in a reduction in value of 60.4% and 66.4%, respectively, if the class is assessed by one individual expert. For comparison, the WoC condition class assessment depicted in Model (c) yields an exponentiated coefficient for Class 2 of 0.579, which results in a 42.1% reduction compared to the reference class. For Classes 3 and 4, value would be reduced by 59.8% and 75.4%, respectively.

This significant difference between the expert's assessment and the WoC assessment emphasizes the importance of property classification in determining market value. It offers a quantifiable measure of the impact of categorization based on the WoC on real estate valuation. Additionally, this comparison illuminates the relative value differences between models that utilize the expert's assessment and those that employ the WoC, highlighting the potential for the WoC to refine and enhance the accuracy of property valuations in markets.

The distinction between the two property condition models, as seen in Table 3 for Model (b) and Model (c), is supported by the Akaike Information Criterion (AIC) for each model.

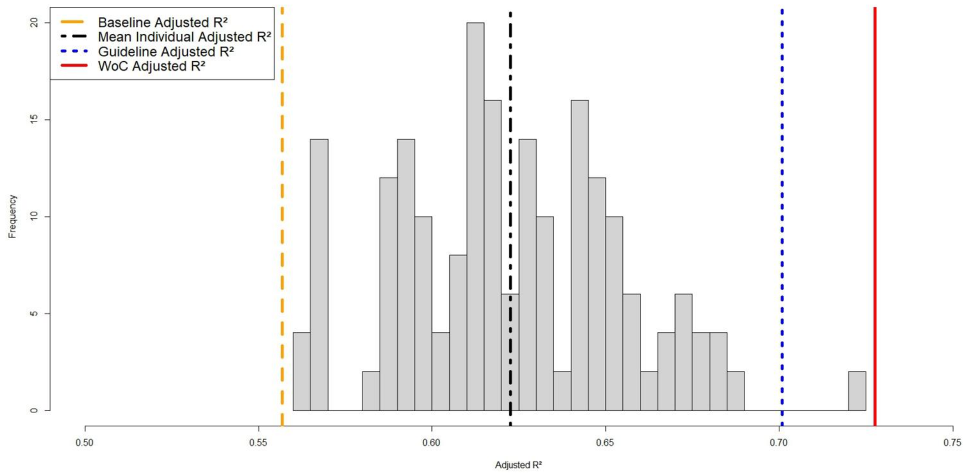


Figure 4. Histogram of the hedonic model adjusted-R-squared
This exhibit displays a histogram of the distribution of adjusted R-squared for all subjects using the hedonic pricing model including the property condition assessment. The mean value of the adjusted R-squared of all individuals is the black dash-dotted vertical line, the guideline adjusted R-squared is dotted blue, and the WoC adjusted R-squared is solid red. The adjusted R-squared of the baseline model is depicted in dashed orange.

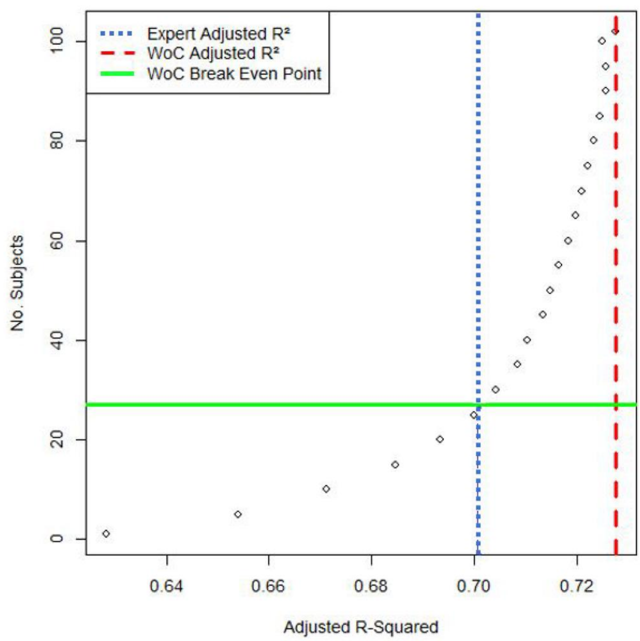


Figure 5. Break-even number of nonprofessionals opinions needed to achieve the professional's level.
This exhibit displays the number of nonprofessional property condition assessments needed to surpass that of a professional appraiser using a Monte Carlo simulation. The red dashed line indicates the WoC adjusted R-squared; the blue dashed line is the R-squared of the professional's model. The horizontal green line shows the 27 subjects in the Monte Carlo simulation that exceed the assessment quality of the professional.

Table 4. ANOVA table for Model 1 (b) and Model 2 (Nested Models).

	Res.Df	RSS	Df	Sum of Sq.	F-value	Pr(>F)	Sign.
Model 1(b)	66	9.55					
Model 2	64	7.69	2	1.86	7.74	0.0010	***

The table shows the F-statistics for the two models including the residual degrees of freedom, the residual sum of squares, the F-statistic, the p-value, and the level of significance of the nested model.

The model that incorporates the WoC assessment achieves a lower AIC (69.57) that the model based on the professional assessment (76.51). This suggests that the hedonic model including the WoC assessment provides a better fit. To assess the significance of the difference between the professional and WoC assessments, we construct a model that includes the real estate professional's assessment along with a variable for the WoC assessment's deviation from the professional assessment. This hedonic model reveals that the variable for positive deviation in the WoC assessment is highly significant, suggesting that a more favorable condition assessment by the WoC notably enhances model quality (detailed in Appendix Table A1). When comparing these two nested models—Model 1 (b) and Model 2—an F-test, as shown in Table 4, confirms that the model incorporating the deviation variable significantly outperforms the other. This evidence underscores the value of including the WoC assessment, particularly when it diverges positively from the professional evaluation, to improve the accuracy and quality of real estate valuation models.

To determine the minimum number of nonprofessional subjects required to surpass the model quality (i.e., adjusted R-squared) achieved by the professional assessment, we conduct a Monte Carlo simulation. This simulation randomly draws 10,000 samples for each N, representing an increasing count of subjects, from the total pool of subjects without replacement. For each subset derived from the Monte Carlo simulation, we calculate the WoC classification for each property. As we increase the number of subjects included in the WoC classification, we use the hedonic model to derive a price estimate that incorporates the simulated WoC variable. Our findings reveal that with an increasing number of nonprofessional subjects, the average adjusted R-squared for the random subsets begins to surpass that of the professional assessment when the number of subjects reaches 27, on average. Figure 4 illustrates this phenomenon, highlighting the diminishing marginal effects as the number of subjects increases. This analysis not only emphasizes the efficacy of the WoC in real estate valuation but also identifies a quantifiable threshold where collective nonprofessional judgments begin to offer superior model quality compared to individual professional assessments.

We also look into the influence of visual manipulations on the assessment of house conditions, as detailed in the previous section. In examining the treatment images, we discover that the presence of a pool in the *Zoom* treatment notably affects the classification by nonprofessionals, skewing the assigned class to a higher category. This finding suggests that positive externalities in the surroundings or neighborhood can bias the evaluation of the property's built quality. However, comparisons between color photos and black and white images, as well as the orientation of the image (whether it is mirrored), do not reveal any significant differences in assessment outcomes. Furthermore, our control images show no biases in the judgment of nonprofessionals at an aggregated level, indicating a consistent evaluation process among participants. Therefore, we conclude that certain visual manipulations and presentation methods can indeed influence the final assessment of property condition based on images, highlighting the need for careful consideration of these factors in property appraisal.

Conclusion

This study leverages a well-documented behavioral economic concept, the wisdom of the crowd (WoC), to enhance the assessment of property condition, thereby refining hedonic pricing models. Our findings indicate that the critical number of nonprofessional individuals required to exceed the expertise of a professional is 27.

Our findings demonstrate that the WoC is more precise at predicting variables that affect real estate prices. This insight is substantial, considering that the primary customer base in the housing market is the general population, not real estate experts. Therefore, it is understandable that property condition assessments for a considerable number of homes are influenced by factors such as the presence of a pool or the appeal of the surroundings. Such deviations actually improve the precision of the prediction model, which is crucial because property prices are ultimately set by the general home-buying public.

Moreover, our findings reveal that aggregating individual, independent condition assessments yields better price predictions within a hedonic model than relying on a professional's singular assessment. This superior performance arises because a professional cannot fully capture the diverse preferences of market participants, a gap the WoC mechanism fills by weighting these various inputs in its aggregation process. It is important to note, however, that the tendency of subjects to overestimate suggests that although the WoC can integrate diverse inputs to generate accurate outcomes, it may also mirror systematic biases such as overoptimism. The practical implications of these findings underscore the potential for inflated property prices due to crowd overestimation, highlighting the necessity for expert oversight and algorithmic adjustments to refine the WoC approach and counteract its limitations. Given that attractiveness significantly hinges on personal preferences, future research should delve into how specific preferences for features such as pools, building materials, or other factors influence bias in property assessment and pricing models.

Given the general public's greater willingness to share their assessments, in contrast to real estate professionals who may intentionally withhold their expert valuations, leveraging a group from the general population presents a significant opportunity for the price discovery process in an increasingly information-rich era. This approach is particularly appropriate for variables that depend heavily on subjective judgments, especially those based on visual inspections in real estate valuation. This trend underscores the value of the WoC in enhancing transparency and accuracy in property valuation, tapping into the collective judgment and preferences of a broader audience.

Based on our findings, we recommend that future researchers explore the use of computer vision and machine learning algorithms to integrate assessments from the home-buying public into automated valuation models (AVMs). Incorporating the condition variable derived from images into AVMs and computer vision predictions necessitates an objective assessment of the pictures to serve as a reliable ground truth. Therefore, the images used for the ground truth must reflect a condition assessment that aligns as closely as possible with market standards. Future studies should examine the effects of using a ground truth assessed by the WoC versus one evaluated by a professional appraiser. This approach could offer an alternative to relying on the subjective and potentially expensive opinions of individual appraisal professionals, but it also improve mass appraisals using image data. Given the prevalent reliance on the judgment of single real estate professionals to derive an "objective" value for properties, it is crucial to scrutinize the objectivity of such appraisals. This scrutiny invites the exploration of alternative methods that could provide a more objective estimate of true market value, leveraging the collective insights and perceptions of the broader market rather than the singular viewpoint of professionals.

In conclusion, our study highlights the effectiveness of the WoC in assessing the condition of homes. By integrating a wide variety of viewpoints, we significantly enhance the coefficient of determination, thereby improving the accuracy of our predictions. Considering the crucial influence of the general prospective home-buying public on real estate prices, leveraging the WoC appears to be a logical progression for accurately predicting factors that impact home values.

Notes

1. This process has been termed a "drive-by appraisal" because appraisers often do not even step foot on the property. And if they do, it has become common to avoid entering the house. As such, even in a post-COVID era, appraisals are regularly completed without firsthand viewing of the interior of the home.
2. Seiler et al. (2012) use ocular tracking to examine the importance of photos in the home search process, while Johnson et al. (2020) and Benefield et al. (2011) document the importance of visual images in enhancing pricing models.
3. This concept is well illustrated by Galton (1907). All participants had private information, did not know one another, and did not share information on the ox before, during, or after posting their estimates of its weight. For these reasons, no herding behavior occurred either. Due to the independence and simultaneous decision-making of group members, participants are prevented from falling for information cascades, which would introduce bias into subsequent decisions (Bikhchandani et al., 1998).
4. For a list of early studies on group intelligence and the wisdom of crowds, see Surowiecki (2004).
5. Renovation bias in real estate refers to the distortion in property price indices resulting from quality changes due to renovations. This occurs when property values increase because of improvements, potentially leading to an overestimation of overall price growth if these changes are not adequately accounted for in time-series comparisons. Such bias can misrepresent market trends and investment performance over longer periods, highlighting the need for precise methodologies that incorporate renovation effects for more accurate property valuations and index calculations. For a recent study on renovation bias, see Mamre and Sommervoll (2022).
6. In reality, it is both infeasible and impractical to assemble a large group of expert appraisers evaluate a specific property or market.
7. In markets with lower transparency, valuations of single-family houses often rely on the cost approach. The commonly utilized hedonic pricing model includes 30–40 variables. Our method estimates the building's value as the sum of the plot value (expressed in thousands) and the value of the building and its components. The professional appraiser assesses component values by considering manufacturing costs, making adjustments for facilities and location (reflected through land value), and accounting for factors such as age, quality, and construction deficiencies.
8. This accounts for the nonlinear relation between the year of construction and depreciation of the home.
9. Initial evidence on the necessary crowd size has been explored in Walter et al. (2022).

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Tiroler Wissenschaftsförderung.

ORCID

Michael J. Seiler http://orcid.org/0000-0001-7071-2474
Simon Thaler http://orcid.org/0000-0001-7743-3451

References

Ahlfeldt, G., & Mastro, A. (2012). Valuing iconic design: Frank Lloyd Wright architecture in Oak Park, Illinois. *Housing Studies*, *27*(8), 1079–1099. https://doi.org/10.1080/02673037.2012.728575

Armstrong, J. S. (1980). The seer-sucker theory: The value of experts in forecasting. *Technology Review*, *82*(7), 16–24.

Aydin, R., & Smith, B. A. (2008). Evidence of the dual nature of property value recovery following. environmental remediation. *Real Estate Economics*, *36*(4), 777–812. https://doi.org/10.1111/j.1540-6229.2008.00230.x

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, *329*(5995), 1081–1085. https://doi.org/10.1126/science.1185718

Benefield, J. D., Cain, C. L., & Johnson, K. H. (2011). On the relationship between property price, time-on-market, and photo depictions in a multiple listing service. *Journal of Real Estate Finance and Economics*, *43*, 401–422. https://doi.org/10.1007/s11146-009-9219-6

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, *12*(3), 151–170. https://doi.org/10.1257/jep.12.3.151

Billings, S. B. (2015). Hedonic amenity valuation and housing renovations. *Real Estate Economics*, *43*(3), 652–682. https://doi.org/10.1111/1540-6229.12093

Bogin, A. N., & Doerner, W. M. (2019). Property renovations and their impact on house price index construction. *Journal of Real Estate Research*, *41*(2), 249–284. https://doi.org/10.1080/10835547.2019.12091526

Brown, G. (2015). Engaging the wisdom of crowds and public judgment for land use planning using public participation geographic information systems. *Australian Planner*, *52*(3), 199–209. https://doi.org/10.1080/07293682.2015.1034147

Budescu, D. V., & Chen, E. (2012). Identifying expertise and using it to extract the wisdom of the crowds. *Management Science*, *61*, 1–37.

Buitelaar, E., & Schilder, F. (2017). The economics of style: Measuring the price effect of neo-traditional architecture in housing. *Real Estate Economics*, *45*(1), 7–27. https://doi.org/10.1111/1540-6229.12137

Chamberlin, E. H. (1948). An experimental imperfect market. *Journal of Political Economy*, *56*(2), 95–108. https://doi.org/10.1086/256654

Cooke, R., & Goossens, L. (2008). TU delft expert judgment data base. *Reliability Engineering & System Safety*, *93*, 657–674. https://doi.org/10.1016/j.ress.2007.03.005

Ding, C., Simons, R., & Baku, E. (2000). The effect of residential investment on nearby property values: Evidence from Cleveland, Ohio. *Journal of Real Estate Research*, *19*(1), 23–48. https://doi.org/10.1080/10835547.2000.12091011

Eriksen, M. D., Fout, H. B., Palim, M., & Rosenblatt E. (2020). Contract price confirmation bias: Evidence from repeat appraisals. *Journal of Real Estate Finance and Economics*, *60*(1), 77–98. https://doi.org/10.1007/s11146-019-09716-w

Francke, M. K., & van de Minne, A. M. (2017). Land, structure and depreciation. *Real Estate Economics*, *45*(2), 415–451. https://doi.org/10.1111/1540-6229.12146
Galton, F. (1907). Vox populi. *Nature*, *75*, 450–451. https://doi.org/10.1038/075450a0
Gillingham, K. T., & Watten, A. (2024). How is rooftop solar capitalized in home prices? *Regional Science and Urban Economics*, 104006. https://doi.org/10.1016/j.regsciurbeco.2024.104006
Goldstein, D. G., McAfee, R. P., & Suri, S. (2014). The wisdom of smaller, smarter crowds. *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, 471–488. https://doi.org/10.1145/2600057.2602886
Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review*, *112*(2), 494. https://doi.org/10.1037/0033-295X.112.2.494
Hebdzyński, M. (2023). Quality information gaps in housing listings: Do words mean the same as pictures? *Journal of Housing and the Built Environment*, *38*(4), 2399–2425. https://doi.org/10.1007/s10901-023-10043-z
Johnson, E. B., Tidwell, A., & Villupuram, S. V. (2020). Valuing curb appeal. *Journal of Real Estate Finance and Economics*, *60*(1), 111–133. https://doi.org/10.1007/s11146-019-09713-z
Juni, M. Z., & Eckstein, M. P. (2017). The wisdom of crowds for visual search. *Proceedings of the National Academy of Sciences*, *114*(21), E4306–E4315. https://doi.org/10.1073/pnas.1610732114
Kao, A. B., Berdahl, A. M., Hartnett, A. T., Lutz, M. J., Bak-Coleman, J. B., Ioannou, C. C., Giam, X., & Couzin, I. D. (2018). Counteracting estimation bias and social influence to improve the wisdom of crowds. *Journal of the Royal Society Interface*, *15*(141), 20180130. https://doi.org/10.1098/rsif.2018.0130
Keynes, J. (1936). *The general theory of employment, interest and money*. Harcourt, Brace.
Knight, J. R., & Sirmans, C. (1996). Depreciation, maintenance, and housing prices. *Journal of Housing Economics*, *5*(4), 369–389. https://doi.org/10.1006/jhec.1996.0019
Koch, D., Thaler, S., & Mayr, R. (2021). Order versus disorder: The impact on value estimates of durable consumption goods. *Applied Economics Letters*, *28*(19), 1635–1640. https://doi.org/10.1080/13504851.2020.1841081
Koriat, A. (2012). When are two heads better than one and why? *Science*, *336*(6079), 360–362. https://doi.org/10.1126/science.1216549
Kostic, Z., & Jevremovic, A. (2020). What image features boost housing market predictions? *IEEE Transactions on Multimedia*, *22*(7), 1904–1916. https://doi.org/10.1109/TMM.2020.2966890
Lakens, D., Fockenberg, D. A., Lemmens, K. P., Ham, J., & Midden, C. J. (2013). Brightness differences influence the evaluation of affective pictures. *Cognition & Emotion*, *27*(7), 1225–1246. https://doi.org/10.1080/02699931.2013.781501
Leguizamon, S. (2010). The influence of reference group house size on house price. *Real Estate Economics*, *38*(3), 507–527. https://doi.org/10.1111/j.1540-6229.2010.00275.x
Lindenthal, T., & Johnson, E. B. (2021). Machine learning, architectural styles and property values. *Journal of Real Estate Finance and Economics*. Advance online publication. https://doi.org/10.1007/s11146-021-09845-1
Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the national academy of sciences*, *108*(22), 9020–9025. https://doi.org/10.1073/pnas.1008636108
Luchtenberg, K. F., Seiler, M. J., & Sun, H. (2019). Listing agent signals: Does a picture paint a thousand words? *Journal of Real Estate Finance and Economics*, *59*, 617–648. https://doi.org/10.1007/s11146-018-9674-z
Maloney, M. T., & Mulherin, J. (2003). The complexity of price discovery in an efficient market: The stock market reaction to the challenger crash. *Journal of Corporate Finance*, *9*(4), 453–479. https://doi.org/10.1016/S0929-1199(02)00055-X
Mamre, M. O., & Sommervoll, D. E. (2022). Coming of age: Renovation premiums in housing markets. *Journal of Real Estate Finance and Economics*. Advance online publication. https://doi.org/10.1007/s11146-022-09917-w
Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S. E., Moore, D., Atanasov, P., Swift, S. A. (2014). Psychological strategies for winning a geopolitical forecasting Tournament. *Psychological Science*, *25*(5), 1106–1115. https://doi.org/10.1177/0956797614524255

Miller, N., Sah, V., & Sklarz, M. (2018). Estimating property condition effect on residential property value: Evidence from U.S. home sales data. *Journal of Real Estate Research*, *40*(2), 179–198. https://doi.org/10.1080/10835547.2018.12091497

Morgan, M. G. (2014). Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academy of Sciences*, *111*, 7176–7184. https://doi.org/10.1073/pnas.1319946111

Murr, A. E. (2011). "Wisdom of crowds"? A decentralised election forecasting model that uses citizens' local expectations. *Electoral Studies*, *30*(4), 771–783. https://doi.org/10.1016/j.electstud.2011.07.005

Nakamura, S., Peiser, R., & Torto, R. (2018). Are there investment premiums for mixed-use properties? *Journal of Real Estate Research*, *40*(1), 1–40. https://doi.org/10.1080/10835547.2018.12091489

Nanda, A., & Ross, S. L. (2012). The impact of property condition disclosure laws on housing prices: Evidence from an event study using propensity scores. *Journal of Real Estate Finance and Economics*, *45*(1), 88–109. https://doi.org/10.1007/s11146-009-9206-y

Orquin, J. L., Perkovic, S., & Grunert, K. G. (2018). Visual biases in decision making. *Applied Economic Perspectives and Policy*, *40*(4), 523–537. https://doi.org/10.1093/aepp/ppy020

Page, S. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton University Press.

Palacios, J., Eichholtz, P., Kok, N., & Aydin, E. (2021). The impact of housing conditions on health outcomes. *Real Estate Economics*, *49*(4), 1172–1200. https://doi.org/10.1111/1540-6229.12317

Palley, A. B., & Soll, J. B. (2019). Extracting the wisdom of crowds when information is shared. *Management Science*, *65*(5), 2291–2309. https://doi.org/10.1287/mnsc.2018.3047

Pavlov, A., & Blazenko, G. W. (2005). The neighborhood effect of real estate maintenance. *Journal of Real Estate Finance and Economics*, *30*(4), 327–340. https://doi.org/10.1007/s11146-005-7011-9

Ray, R. (2006). Prediction markets and the financial "Wisdom of Crowds." *Journal of Behavioral Finance*, *7*(1), 2–4. https://doi.org/10.1207/s15427579jpfm0701_1

Ready, R. (2010). Do landfills always depress nearby property values? *Journal of Real Estate Research*, *32*(3), 321–340. https://doi.org/10.1080/10835547.2010.12091279

Schelling, T. C. (1980). *The strategy of conflict: With a new preface by the author*. Harvard University Press.

Seiler, M., Madhavan, P., & Liechty, M. (2012). Toward an understanding of real estate homebuyer internet search behavior: An application of ocular tracking technology. *Journal of Real Estate Research*, *34*(2), 211–242. https://doi.org/10.1080/10835547.2012.12091333

Shanteau, J. (1995). Expert judgment and financial decision making. In B. Green (Ed.), *Risk behaviour and risk management: Proceedings of the First International Stockholm Seminar on Risk Behaviour and Risk Management* (pp. 16–32).

Simmons-Mosley, T. (2003). Interdependence effects of housing abandonment and renovation. *Journal of Real Estate Research*, *25*(4), 421–430. https://doi.org/10.1080/10835547.2003.12091122

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99–118. https://doi.org/10.2307/1884852

Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of Political Economy*, *70*(2), 111–137. https://doi.org/10.1086/258609

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. Doubleday.

Thaler, S., & Koch, D. (2023). Real estate pictures: The role of furniture preferences in subjective valuation. *Journal of Housing Research*, 32(2), 180–203. https://doi.org/10.1080/10527001.2023.2168585

Turnbull, G. K., Dombrow, J., & Sirmans, C. F. (2006). Big house, little house: Relative size and value. *Real Estate Economics*, *34*(3), 439–456. https://doi.org/10.1111/j.1540-6229.2006.00173.x

Turner, T. M., & Seo, Y. (2021). House prices, open space, and household characteristics. *Journal of Real Estate Research*, *43*(2), 204–225. https://doi.org/10.1080/08965803.2021.1925498

van Dolder, D., & van den Assem, M. J. (2018). The wisdom of the inner crowd in three large natural experiments. *Nature Human Behaviour*, *2*(1), 21–26. https://doi.org/10.1038/s41562-017-0247-6

Voicu, I., & Been, V. (2008). The effect of community gardens on neighboring property values. *Real Estate Economics*, *36*(2), 241–283. https://doi.org/10.1111/j.1540-6229.2008.00213.x

Walter, V., Kölle, M., & Collmar, D. (2022). Measuring the wisdom of the crowd: How many is enough? *Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, *90*(3), 269–291.

Wang, X., Takada, Y., Kado, Y., & Yamasaki, T. (2019). Predicting the attractiveness of real estate images by pairwise comparison using deep learning. In *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 84–89). IEEE. https://doi.org/10.1109/ICMEW.2019.0-106

Wiley, J. A., & Zumpano, L. V. (2008). The real estate brokerage market and the decision to disclose property condition defects. *Real Estate Economics*, *36*(4), 693–716. https://doi.org/10.1111/j.1540-6229.2008.00227.x

Wilhelmsson, M. (2008). House price depreciation rates and level of maintenance. *Journal of Housing Economics*, *17*(1), 88–101. https://doi.org/10.1016/j.jhe.2007.09.001

Winkielman, P., & Gogolushko, Y. (2018). Influence of suboptimally and optimally presented affective pictures and words on consumption-related behavior. *Frontiers in Psychology*, *8*, 2261. https://doi.org/10.3389/fpsyg.2017.02261

Yi, S. K. M., Steyvers, M., Lee, M. D., & Dry, M. J. (2012). The wisdom of the crowd in combinatorial problems. *Cognitive Science*, *36*(3), 452–470. https://doi.org/10.1111/j.1551-6709.2011.01223.x

Zabel, J. E. (1999). Controlling for quality in house price indices. *Journal of Real Estate Finance and Economics*, *19*(3), 223–241. https://doi.org/10.1023/A:1007832527244

Appendix

Figure A1. Screenshot of the Example Comparison Task.
This exhibit shows sample images seen by the subjects when tasked with assessing the 100 homes in our sample across four different classes of property condition.

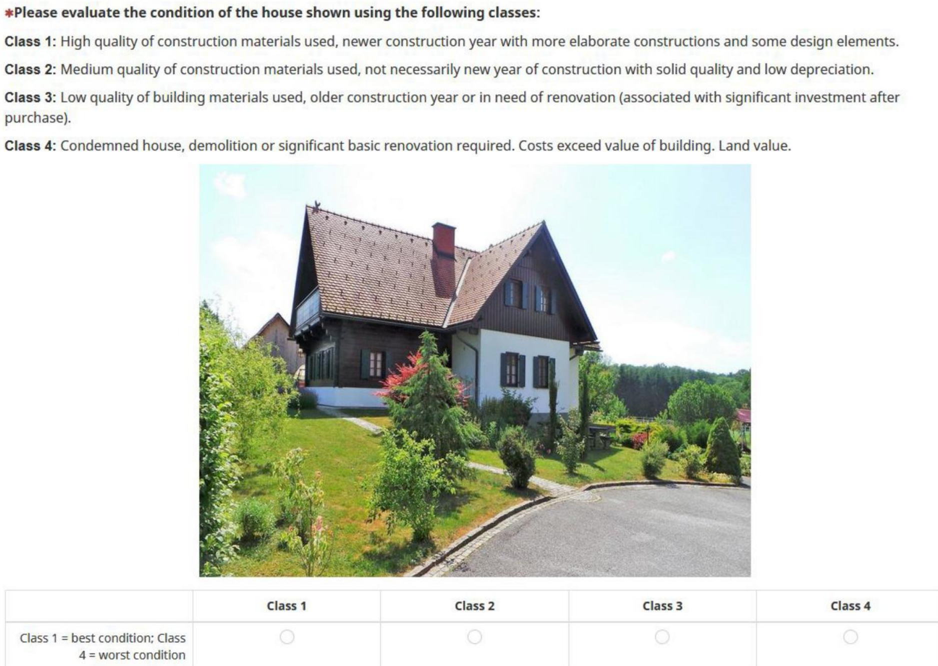


Figure A2. Screenshot of the Condition Evaluation Task for One Sample Home.
This exhibit shows what participants see when asked to evaluate the condition of the home.

Table A1. Regression models reflecting deviations between professionals and nonprofessionals.

	Dependent Variable: log(Price)
LAND VALUE [1K]	0.002***
	(0.0002)
SQUARE FOOTAGE	0.003***
	(0.001)
poly(YEAR, 2)1	−0.091
	(0.465)
poly(YEAR, 2)2	−0.593
	(0.388)
EXP_CLASS 2	−0.556***
	(0.156)
EXP_CLASS 3	−1.185***
	(0.181)
EXP_CLASS 4	−1.444***
	(0.222)
CLASS_DEV −1	−0.102
	(0.209)
CLASS_DEV 1	0.393***
	(0.102)
Constant	12.982***
	(0.207)
Observations	74
R^2	0.782
Adjusted R^2	0.752
Residual Std. Error	0.347 (df = 64)
F Statistic	25.545*** (df = 9; 64)

Note: *p <0.1; **p <0.05; ***p <0.01
This exhibit reports results from the regression including the professional condition assessment variable and a variable of the deviation of the WoC assessment to the professional assessment together with the baseline hedonic model.

Table A2. Regression Price of All Homes Excluding the Year of Construction.

	Dependent Variable:		
	log(price)		
	(1)	(2)	(3)
LAND VALUE [1K]	0.002***	0.002***	0.002***
	(0.0003)	(0.0003)	(0.0003)
SQUARE FOOTAGE	0.003***	0.003***	0.002***
	(0.001)	(0.001)	(0.001)
EXP_CLASS 2		−0.310**	
		(0.153)	
EXP_CLASS 3		−0.795***	
		(0.152)	
EXP_CLASS 4		−1.007***	
		(0.172)	
WOC_CLASS 2			−0.476***
			(0.127)
WOC_CLASS 3			−0.865***
			(0.133)
WOC_CLASS 4			−1.319***
			(0.180)
Constant	12.084***	12.776***	12.854***
	(0.141)	(0.178)	(0.155)
Observations	92	92	92
R^2	0.439	0.664	0.692
Adjusted R^2	0.427	0.645	0.674
Residual Std. Error	0.515 (df = 89)	0.405 (df = 86)	0.388 (df = 86)
F statistic	34.865***	34.037***	38.613***

This exhibit repeats the regression results of the models with the year of construction removed.