When houses wear faces: Reverse correlation applied to architectural design

Kira Pohlmann a,⁎, Nour Tawil a, Timothy R. Brick b, Simone Kühn a,c,d

a Max Planck Institute for Human Development, Center for Environmental Neuroscience, Lentzeallee 94, 14195, Berlin, Germany
b Pennsylvania State University, Department of Human Development and Family Studies, University Park, PA, USA
Cc Clinic and Polyclinic for Psychiatry and Psychotherapy, University Medical Center Hamburg- Eppendorf, Martinistr. 52, 20251, Hamburg, Germany
打卡Max Planck UCL Centre for Computational Psychiatry and Ageing Research Berlin, Germany and London, UK, Lentzeallee 94, 14195, Berlin, Germany

A R T I C L E   I N F O
Handling Editor: L. McCunn
Keywords:
Reverse correlation
Architecture
Housing exterior
House facades
Pareidolia

A B S T R A C T
Reverse correlation (RC) is a data-driven method from social psychology that has been effectively shown to visualize the mental representations that humans hold regarding facial attributes. The method helps to understand what features are relevant in terms of the evaluation of faces, such as dominance or submissiveness. To the best of our knowledge, RC has solely been applied to faces within the area of psychology until this point. However, there are many other areas where it is of interest to understand how humans evaluate and visualize content, one of them being the evaluation of house facades. With this work, we extended the application of RC to architectural design, specifically focusing on the evaluation of house facades with respect to the psychological attributes of facelikeness, invitingness, and likeability. Furthermore, we propose a novel approach to create the base image, by utilizing a generative adversarial network. In an online study with a between-subject design, 121 participants completed the RC task, with 40 to 41 participants assigned to each of the three attributes. The resulting classification images (CIs) from the RC task unveil face-related features for the attribute facelikeness, signifying the potential extension of the RC methodology beyond the established domain of facial analysis to other domains, such as architectural design.

1. Introduction

The data-driven method of reverse correlation (RC) (Dotsch et al., 2008) has gained popularity in social psychology in recent years (Mangini & Biederman, 2004; Brinkman, Todorov, & Dotsch, 2017; Kevane & Koopmann-Holm, 2021; Schmitz, Rougier, & Yzerbyt, 2021). This method offers a unique approach to visualize mental representations and subjective perceptions of faces, eliminating the necessity to explicitly articulate the features affecting the perception (Dotsch & Todorov, 2012; Brinkman, Todorov, & Dotsch, 2017).

In previous studies, the RC method has effectively visualized features related to the perception of race, gender, and personality traits, such as dominance or submissiveness (Dotsch & Todorov, 2012; Brinkman, Todorov, & Dotsch, 2017; Imhoff & Dotsch, 2013; Imhoff et al., 2011; Oliveira, Garcia-Marques, & Dotsch, 2019; Poveda-Bautista, Diego-Mas, & Alcaide-Marzal, 2021). The procedure involves overlaying a grayscale base image with random noise. The base image typically consists of a mean face, which is a face averaged across multiple images of faces (Dotsch et al., 2008). Participants are then presented with pairs of the noise-stimuli, and in each trial, they choose the stimulus that best represents or aligns with a specific attribute (Dotsch et al., 2008). The subsequently selected images are averaged to form the classification image (CI) (Dotsch et al., 2008). Notably, regions including the mouth, eyes, eyebrows, and hair have consistently emerged as relevant in the context of social face perception (Dotsch & Todorov, 2012).

Identifying the relevant areas within a CI and distinguishing between CIs containing meaningful information and those that do not is one of the challenges of RC. Commonly, the CIs are rated on the same attributes by an independent sample of participants (Brown-Iannuzzi et al., 2017; Cone et al., 2021; Oliveira, Garcia-Marques, & Dotsch, 2019). An alternative method is using the informational value (infoVal) metric (Brinkman et al., 2019b). This metric employs a scoring system to quantify the likelihood of a CI containing information beyond random chance. By using the infoVal metric, researchers gain a quantitative measure that supports the evaluation of the CIs, as it determines which CIs contain valuable information.

⁎ Corresponding author.
E-mail address: pohlmann@mpib-berlin.mpg.de (K. Pohlmann).

https://doi.org/10.1016/j.jenvp.2024.102401
Received 2 January 2024; Received in revised form 25 June 2024; Accepted 7 August 2024
Available online 8 August 2024
0272-4944/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).
The predominant version of the RC task is the two-images forced choice (2IFC) task (Imhoff et al., 2011; Brinkman et al., 2019a; Brown-Iannuzzi et al., 2017; Dotsch & Todorov, 2012; Imhoff & Dotsch, 2013; Imhoff & Messer, 2019; Imhoff et al., 2013; Oliveira, Garcia-Marques, & Dotsch, 2019). This task presents participants with one image pair at a time, each consisting of the base image superimposed with oriented noise (random noise) and the base image superimposed with inverse noise (negative oriented noise). The noise patterns differ for each trial, such that different features of the base image are highlighted by the darker or lighter parts of the noise throughout the RC task. The participant’s objective is to select the image that best aligns with the given attribute, e.g., dominance (Dotsch & Todorov, 2012). Alternatives are the four-alternatives forced choice (4AFC) task (Brinkman, Todorov, & Dotsch, 2017) or the Brief-RC task (Schmitz, Rougier, & Yzerbyt, 2021).

In the 4AFC task, participants rate a single image overlaid with noise on a 4-point scale, and these ratings serve as weights in the calculation of the CIs (Brinkman, Todorov, & Dotsch, 2017). On the other hand, the Brief-RC task involves presenting participants with a larger number of stimulus pairs, typically six pairs, and requiring them to choose the image that best fits the specified attribute (Schmitz, Rougier, & Yzerbyt, 2021).

To date, the RC method has predominantly found application in the domain of faces, primarily within the confines of social psychology. However, the utility of the RC method beyond facial perception in diverse areas within the broader field of psychology has not yet been explored. This method offers great potential that could prove beneficial in various psychological domains. The current project made an effort to apply the RC task in a novel context — the evaluation of house facades within the realm of environmental psychology. This aims to bridge the gap between the traditionally face-centric applications of RC and the potential insights that can be gained when exploring non-social stimuli. This project sought to explore the capacities of the RC method in understanding perceptual and evaluative processes in the context of architectural aesthetics and environmental psychology.

Given the substantial amount of time we spend in indoor environments or in proximity to buildings (Evans & McCoy, 1998), gaining a better understanding of how individuals perceive and evaluate these structures is of high relevance. The assessment of our immediate physical surroundings, particularly architectural design, involves a multifaceted consideration of various factors. On the one hand, buildings possess the capacity to elicit a spectrum of emotional responses, ranging from stress to relaxation (Ulrich et al., 1991). The phenomenon of pariedolia, wherein faces are perceived in everyday objects (Zhou & Meng, 2020), extends to house facades. Numerous studies have highlighted that detecting a face in a building’s facade can evoke emotional responses in viewers (Chalup, Hong & Ostwald, 2008, 2016; Abbas & Chalup, 2021). On the other hand, a large number of psychological attributes play an important role in the assessment and characterization of architectural design. These attributes include hominess (Coburn et al., 2020), invitingness (Graham, Gosling, & Travis, 2015; Roessler et al., 2022), safety (Appleton, 1996), liking (Roessler et al., 2022), and familiarity (Roessler et al., 2022). It is noteworthy that design features, including colours (Chatterjee & Vartanian, 2014), contours (Tawil, Ascone, & Kühn, 2022; Vartanian et al., 2013), and lighting (Amundottir et al., 2017), have been empirically demonstrated as significant contributors to human responses to architectural design. A comprehensive understanding and analysis of these diverse aspects is crucial for unraveling how individuals interact with and evaluate the built environment. This not only sheds light on the emotional and psychological dimensions of architectural experiences but also lays the foundation for informed and purposeful design practices.

To explore the applicability of the RC task beyond its conventional use in social psychology, we implemented it on house facades. In an online experiment, participants engaged in the RC task, focusing on the psychological attributes of facelikeness, invitingness, and likeability—attributes closely linked to the assessment of house facades. With this study, we aimed to examine the feasibility of employing the RC task in a non-social psychology context and with that to better understand how individuals assess and attribute psychological characteristics to architectural elements.

2. Materials and methods

In the following sections, we describe the process of stimulus selection and creation for the RC task. We propose a new method to generate the base image of the RC task by utilizing a generative adversarial network (GAN). Lastly, we outline the conducted online experiment in detail, including the experimental design and procedures.

2.1. Stimuli

2.1.1. Base image

A fundamental component of the RC task is the creation of the base image. In the conventional domain of faces, this is achieved by averaging across multiple face images (Dotsch et al., 2008). However, when applying the RC task to houses, replicating the same approach proved challenging due to the greater variability in the features inherent to houses. Unlike faces, houses exhibit a broader range of architectural features, including porches, doors, windows, roofs, and related facade structures. Questions such as the typical position of a door, the number of windows, and their distribution on a facade add to the complexity of generating an average house image. Further, the uniformity of facial outlines (oval shape) demonstrates greater consistency than that observed in houses. The increased complexity presented by houses, in contrast to the relatively standardized features of faces, poses a unique challenge in selecting the base image for the RC task in the domain of architectural stimuli.

To address the challenge of creating a base image for the RC task with house facades, we opted for a novel methodology. Instead of relying on the traditional approach of averaging over a set of images, we leveraged a generative adversarial network (GAN). A GAN consists of two interconnected networks: a generator, responsible for generating synthetic images from input vectors, and a discriminator, tasked with distinguishing between fake images produced by the generator and real images from the training dataset (Goodfellow et al., 2014).

Throughout the training process, these networks engage in an adversarial dynamic, with the generator getting better in its ability to produce images resembling the training dataset, and the discriminator improving its capacity to discern between real and generated images (Goodfellow et al., 2014). This adversarial process helps both networks enhance their capabilities, leading to the generator producing increasingly realistic images over time. Concurrently, the generator learns to map vectors from a so-called latent space into realistic artificial representations. The latent space is a high-dimensional space where each vector corresponds to a potential image. Once the training process is completed, the generator operates independently to create new images from random input vectors. These generated images resemble the training images but are not exact copies; instead, they are unique creations based on the patterns learned during training. For more information, see Goodfellow et al. (2014).

A technique called the truncation trick (Marchesi, 2017) allows for control over this latent space by adjusting a parameter known as the truncation factor \( \psi \). Setting \( \psi = 0.0 \) maximally restricts the latent space, resulting in a singular image that encapsulates the average representation of all vectors within the space (Karras et al., 2020). In contrast, by increasing \( \psi \), the variety and distinctiveness between the generated images can be increased (Karras et al., 2020).

We trained StyleGAN2-ADA (Karras et al., 2020), a state-of-the-art GAN designed for smaller training datasets, on the CalHouses dataset (references to the used images can be found at https://github.com/k-pohlmann/artificial-house-facades). The CalHouses dataset, comprising 2000 images of detached Californian houses, was curated.
from the Houses dataset (Ahmed & Moustafa, 2016) and the SoCal2 dataset (https://www.kaggle.com/datasets/ted8080/house-prices-nd-images-socal, last accessed June 2022). Given that training a GAN requires a dataset of a certain size, typically more than 1000 images (Karras et al., 2020), and due to the lack of suitable datasets containing a diverse range of houses from various countries, including, e.g., Germany, we opted to use the CalHouses dataset. The challenge lies not merely in finding images of houses, but in curating datasets of appropriate size with high-quality images, privacy-compliant content, and known locations of the houses.

We trained StyleGAN2-ADA using the official implementation for PyTorch (https://github.com/NVlabs/stylegan2-ada-pytorch, last accessed Nov 2022). We trained the model from scratch and for approximately 10M iterations (10 million real images shown to the discriminator). Further, the following model and training parameters were used to train StyleGAN2-ADA: GPUs = 1, mirrorx = true, mirrory = false, lrate = 0.002, gamma = 50, batch = 4, β1 = 0, β2 = 0.99, ε = 1 x 10^-8. The final snapshot had a Fréchet inception distance (FID) (Heusel et al., 2017) of 21.02 and a kernel inception distance (KID) (Brinkowski et al., 2018) of 0.0115. We applied the truncation trick (Marchesi, 2017) to control the distribution of the latent space of the GAN. Setting the truncation factor to ω = 0.0 allowed us to obtain an average image that holds the central tendencies of the latent space.

The preprocessing of the base image involved several steps implemented in the Python programming language. First, we resized the base image to dimensions of 512 x 512 pixels. Afterwards, we applied Gaussian blur to the base image using the function ImageFilter.GaussianBlur(radius=10.0) from the Pillow package (https://python-pillow.org/, version 10.1.0, last accessed Oct 2023). The resulting image was then saved in grayscale. To determine the optimal radius for the Gaussian blur, we conducted tests with various radii and selected the one that produced an output similar to the face base images, as observed in previous studies (Imhoff et al., 2019). The processed base image is displayed in Fig. 1. The decision to use a GAN introduces a dynamic and adaptive element to the creation of base images, capturing the variability in architectural features across different houses.

2.1.2. Stimulus pairs

For the RC task, we opted for the 2IFC (two-images forced choice) paradigm, where participants are presented with pairs of images side by side, and they are instructed to select one of the images. The 2IFC paradigm is the most commonly used RC task and proved to be effective for various attributes (Imhoff et al., 2011; Brinkman et al., 2019a; Brown-Iannuzzi et al., 2017; Dotsch & Todorov, 2012; Imhoff & Dotsch, 2013; Imhoff & Messer, 2019; Imhoff et al., 2013; Oliveira, Garcia-Marques, & Dotsch, 2019). The stimulus set was generated in RStudio (Version 2020.02.2) using the R package rcicr (https://rdocumentation.org/packages/rcicr/versions/0.3.4.1, last accessed Oct 2023). Specifically, we utilized the function generateStimuli2IFC() to create 300 stimulus pairs. Each stimulus pair consisted of the base image overlaid with noise (stimulus 1) and a counterpart with inverse noise (stimulus 2). The first stimulus was created by superimposing random noise onto the base image (Dotsch & Todorov, 2012). This noise pattern inherently highlights random regions of the image due to the varying intensities of the noise. Consequently, the noise pattern modifies the appearance of the base image. The second stimulus in each pair was generated by inverting the initial noise pattern, resulting in previously darker areas becoming brighter and vice versa (Dotsch & Todorov, 2012). An example of a stimulus pair is illustrated in Fig. 1, showcasing the base image overlaid with noise and its counterpart with inverse noise. The decision to conduct 300 trials aligns with the typical range observed in RC tasks, which generally spans from 300 to 1000 trials (Brinkman, Todorov, & Dotsch, 2017). Importantly, insights from information value analyses (Brinkman et al., 2019b) informed our decision to keep the task concise at 300 trials, as it showed that more trials do not lead to more signal in the resulting CIs. This choice aims to maintain participants’ engagement and prevent random responses.

2.1.3. Attributes

There are several attributes that are relevant in terms of how architectural design is perceived and evaluated. We decided to include three attributes in this study to initially apply RC to house facades. The first attribute chosen is facelikeness, a relevant evaluation criterion concerning house facades (Roessler et al., 2022; Filliter et al., 2016). This choice is further grounded in the phenomenon of face pareidolia (Zhou & Meng, 2020), where individuals tend to perceive faces in everyday objects, including houses. This tendency is linked to emotional responses, and the recognition of a facial pattern in a house or object can evoke such responses (Chalup, Hong, & Ostwald, 2008, 2010; Abbas & Chalup, 2021; Sussman & Hollander, 2021). Various approaches in the literature have focused on locating facial patterns on house facades (Chalup, Hong, & Ostwald, 2010; Hong, Chalup, & King, 2014). The inclusion of facelikeness as an attribute in our study aims to capture the psychological association with perceiving houses as having facial features, providing valuable insights into the emotional and evaluative aspects of architectural experiences.

Further, we decided to incorporate the attribute of invitingness into the RC task. This attribute is one of the main characteristics in describing houses and plays an important role when individuals are seeking potential homes (Graham, Gosling, & Travis, 2015). Additionally, we
opted to include likeability as a third attribute in our study as this enables an exploration of how participants visualize houses, they find appealing. This attribute has demonstrated significance in the context of rating and evaluating house facades, with a positive correlation observed between likeability and typicality ratings (Roessler et al., 2022; Filliter et al., 2016).

Both invitingness and likeability are relevant attributes in the evaluation of house facades (Graham, Gosling, & Travis, 2015; Roessler et al., 2022). However, unlike facelikeness, they are not associated with specific human facial features, such as eyes and mouth. To investigate how mental representations differ between these two types of attributes—facelikeness being strictly related to specific features and invitingness and likeability being more abstract—we included all three attributes in our study. Specifically, we aimed to determine whether different features would be highlighted in the RC task, potentially connecting specific characteristics to particular attributes, such as the entrance door for invitingness.

This approach enables us to gain insights into how participants perceive and attribute psychological characteristics to architectural elements.

2.2. Participants

The RC task was implemented as part of an online experiment hosted on the Prolific platform (https://www.prolific.co/, last accessed Oct. 2023). The experiment was designed as a between-subject study, with the goal of recruiting 40 participants for each of the three attributes. The inclusion of 40 participants in our study aligns with the established literature in the field of RC (Dotsch et al., 2008; Brinkman et al., 2019a; Imhoff & Dotsch, 2013). In total, 123 participants completed the online experiment. To ensure the quality of responses, participants who completed the RC task in less than 2 min were excluded from the analysis. It was likely that participants responded randomly to the stimuli pairs when finishing the task in such a short time frame. All participants were of legal age and based in Germany. We decided to include German participants, despite the base image showing a Californian house. This decision aimed to avoid any bias among participants from the United States regarding the specific type of house depicted. The base image might be more recognizable to some participants from the United States depending on their geographic location within the country, whereas it is likely to be unfamiliar to the German participants. Additionally, we took measures to have an equal distribution of male and female participants for each attribute. At the beginning of the experiment, participants were asked to confirm they did not have any knowledge and interest in architecture and design (Tawil et al., 2023), or overthinking the decision-making process. The RC task started with a brief practice run comprising five trials and aimed to familiarize participants with the task dynamic. Following this, participants were briefed that the main task was about to start and entailed 300 trials. In each trial, a stimulus pair was presented on the screen, together with the instructions to select the house that most closely matched the assigned attribute (facelike, inviting, or likeable). Participants engaged in the RC task in a continuous session without a break.

After the RC task, participants were prompted to provide general demographic information and answer questions about their current living situation. This included, e.g., questions about the housing type (detached house or apartment complex), the size of the participants’ living space, and whether they reside alone or with others. Further, they were requested to indicate the average number of days per week they spent at home. Additionally, participants were asked to respond to an adapted German version of the Vienna Art Interest and Art Knowledge Questionnaire (VAIAK) (Speckter et al., 2020) modified to reflect knowledge and interest in architecture and design (Tawil et al., 2023), as well as the Individual Differences in Anthropomorphism Questionnaire (IDAQ) translated into German (Waytz, Cacioppo, & Epley, 2010), and the Big Five Inventory-SOEP (BFI-S) (Schupp & Gerlitz, 2008).

3. Results

The study was pre-registered and can be found at https://aspredicted.org/7xk3s.pdf. In the initial phase of analysis, we generated CIs grouped by the attribute (attribute CIs), resulting in one CI for each of the three attributes: facelikeness, invitingness, and likeability. A CI is constructed by averaging all images selected by participants during the RC task (Dotsch & Todorov, 2012). To compute the CIs, we employed the R package rcicr and utilized the function batchGenerateCIZIFC(). The resulting attribute CIs are depicted in Fig. 2.

Inspired by the cluster tests performed on the CIs in previous studies (Dotsch & Todorov, 2012), we aimed to visualize the most distinct areas of the CIs. However, as the CIs in our study were not rated by an individual sample of participants as in some of the previous works (Dotsch & Todorov, 2012), we opted for a simpler approach to visualize clusters of features. As described in prior research (Dotsch & Todorov, 2012), we smoothed the CIs using a Gaussian filter with \( \sigma = 4 \) (ImageFilter.GaussianBlur(radius = \( \sigma \)) from the Pillow package https://python-pillow.org/, version 10.1.0, last accessed Oct 2023). Afterwards, the pixel values of the CI were normalized using a Z-transformation. Next, a two-tailed thresholding operation was performed on the Z-transformed image to identify clusters of pixels. In this context, a cluster is defined as...
a group of pixels where the absolute Z-score was greater than or equal to a critical value of 2.3. This allowed us to identify clusters of pixels that significantly differed from the overall image. The clusters are shown in Fig. 3, overlayed on the base image to highlight the relevant differences and the related architectural features.

Following the generation of CIs grouped by attributes, we generated individual CIs for each participant, which we used for an infoVal analysis (Brinkman et al., 2019b). The infoVal metric assesses whether a CI contains more than random information (Brinkman et al., 2019b), offering a quantitative measure to complement the visual analysis of the CIs. This analysis allowed us to quantify the number of CIs that contain meaningful information for each attribute, ensuring a focus on relevant features. For the infoVal analysis, we utilized the R package rcr and employed the function computeInfoVal2IFC(). CIs with an infoVal score above 1.96 are considered to contain meaningful information (Brinkman et al., 2019b). In the case of facelikeness, 12 out of the 41 CIs showed a score above the critical value, ranging from 2.034 to 6.749. For invitingness, eight out of 40 CIs surpassed the critical value, with infoVal scores ranging from 2.089 to 6.567. Lastly, for likeability, six out of the 40 CIs had a score above the critical value, ranging from 2.210 to 6.820. All CIs with an infoVal score above 1.96 are presented in Fig. 4. Fig. 5 provides an overview of all calculated infoVal scores.

Further, we calculated summary scores for the VAIAK part A, the subscale revealing the interest in art and architecture, and the IDAQ, assessing individual differences in anthropomorphism (scores can be found in Table 1). Shapiro-Wilk tests indicated that the VAIAK and IDAQ summary scores were normally distributed within the groups (facelikeness, invitingness, and likeability) and Levene’s test of homogeneity of variances showed no significant differences. A one-way ANOVA was conducted for the summary scores as the dependent variable and the groups as the independent variable. The one-way ANOVA was not significant for the VAIAK summary scores (F(2,118) = 1.787, p = 0.172) and the IDAQ summary scores (F(2,118) = 0.618, p = 0.541).

A correlation was conducted to assess the relationship between VAIAK (part A) summary scores and infoVal scores, as well as IDAQ summary scores and infoVal scores for each of the three groups. Shapiro-Wilk tests indicated that the infoVal scores were not normally distributed. A non-parametric Spearman’s rank-order correlation indicated no significant correlation between VAIAK summary scores and infoVal scores for facelikeness (rs = −0.021, p = 0.895), invitingness (rs = 0.007, p = 0.968), and likeability (rs = −0.010, p = 0.953). A Spearman’s rank-order correlation was not significant between IDAQ summary scores and infoVal scores for facelikeness (rs = 0.091, p = 0.570), invitingness (rs = 0.086, p = 0.596), and likeability (rs = −0.165, p = 0.309). All tests were conducted using the IBM SPSS Statistics (Version 26).

To investigate whether the participants’ current living situation affected the infoVal score, we conducted the non-parametric Wilcoxon rank-sum test. We included the participants’ current housing type (house or apartment building), previous housing types (whether participants have ever lived in a house), and future plans for housing types (whether participants plan to live in a house in the future). For the latter, participants currently living in a house were grouped with those who intend to live in a house. The test indicated no significant difference in infoVal scores and current housing type for facelikeness (W = 159, p = 0.533), invitingness (W = 129, p = 0.750), and likeability (W = 159, p = 0.793). Similarly, for participants’ previous housing types, the Wilcoxon rank-sum test was not significant for infoVal scores for facelikeness (W = 156, p = 0.061), invitingness (W = 130, p = 0.548), and likeability (W = 134, p = 0.633). Further, the test showed no significant difference in infoVal scores regarding participants’ future plans for housing types for facelikeness (W = 129, p = 0.653), invitingness (W = 149, p = 0.765), and likeability (W = 146, p = 0.698).

Previous studies showed sex differences concerning aesthetic preference and architectural design (Munroe, Munroe, & Lansky, 1976; Tavil, Ascone, & Kühn, 2022). Therefore, we tested for sex differences and infoVal scores within each group, despite maintaining a balanced distribution of sex assigned at birth. We conducted the non-parametric Wilcoxon rank-sum test. The test indicated no significant differences
in InfoVal scores and sex assigned at birth for *facelikeness* \( (W = 238, p = 0.477) \), *invitingness* \( (W = 132, p = 0.069) \), and *likeability* \( (W = 262, p = 0.093) \). The statistical test was run in RStudio using the function `wilcox.test()` of the package `stats` (https://rdocumentation.org/packages/stats/versions/3.6.2, last accessed Nov 2023).

### 4. Discussion

To the best of our knowledge, this is the first project that extends the application of RC in the field of psychology from the domain of faces to house facades. We shifted the focus to exploring the mental representations individuals hold regarding house facades using the RC task. The creation of a base image involved employing the truncation trick and generating an average image (with \( \psi = 0.0 \)) by training StyleGAN2-ADA on the CalHouses dataset. 40 to 41 participants completed the RC task on each of the attributes *facelikeness*, *invitingness*, and *likeability* (\( N = 121 \) in total). We computed CIs averages across the attributes, as well as on a participant level. Additionally, infoVal scores were calculated to ascertain whether individual CIs contained more than random noise.

Examining the attribute CIs, particularly for *facelikeness* (see Fig. 2), noteworthy changes became evident, marking it as the attribute with the strongest alterations when compared to the base image. The CI showcases typical facial attributes superimposed on the house facade, which also reflects on the most distinct pixel clusters (see Fig. 3). Notably, the windows exhibit a darker hue, resembling eyes. A subtle artefact below the windows in the base image, previously inconspicuous, now takes on a more pronounced shape, forming the representation of a potential nose. Even the small window row of the garage door gains clarity, assuming the appearance of a mouth for the face. However, the mouth was not highlighted in the pixel clusters, which indicates that the focus

![Facelikeness](image1.png)

![Invitingness](image2.png)

![Likeability](image3.png)

*Fig. 4.* All classification images (CIs) that had an infoVal score above the critical value 1.96 and contain more than random information.
was rather on the eyes and nose than on the mouth.

In addition to the attribute CIs, we conducted an analysis of participants’ individual CIs and calculated infoVal scores, where a score above 1.96 signified that a CI contained more than random noise and held informational value (Brinkman et al., 2019b). When examining these individual CIs (see Fig. 4), it became evident that the highlighted features of the eyes (windows) and nose (artefact) in the attribute CI manifested consistently across most individual CIs. Notably, certain CIs accentuated the smaller windows of the door more prominently. This observation found additional support in the analysis of pixel clusters (see Fig. 3), as the two clusters on the door window of the base image became one larger cluster in the facelikeness CI. This indicates that there is more than one possible face on the house facade. The highlighted face-related features in the CIs align with the pareidolia theory concerning house facades (Chalup, Hong, & Ostwald, 2008, 2010; Sussman & Hollander, 2021), indicating that participants successfully projected a face-like pattern onto the base image. The attribute CI distinctly revealed the features of the eyes and nose, providing a visual representation of the participants’ mental associations with the concept of facelikeness in the context of house facades.

In contrast to facelikeness, the attribute CIs for invitingness and likeability showed no clear differences from the base image (see Fig. 2). It appears that the RC task did not lead to a visualization of features specifically associated with the attributes of invitingness and likeability. This was further supported by the highlighted pixel clusters (see Fig. 3), since the CIs for invitingness and likeability showed the same clusters as the base image. However, in certain individual CIs for invitingness and likeability (see Fig. 4), distinct variations in the door were observed, particularly in the intensity of its openings. Some CIs depicted darker door openings, while others presented a brighter appearance, making them barely noticeable. This observation suggests that participants hold diverse mental representations of what constitutes an inviting or likeable house. Unlike the evaluation of faces, which is a fundamental aspect of everyday life, the conceptualization and assessment of inviting and likeable architectural features might be more subjective and less universally defined. It seems that, in the context of invitingness and likeability, the door itself is a more relevant feature for these attributes.

The lack of observable changes in the attribute CIs for invitingness and likeability may be explained by the contrasting patterns (dark vs. light door) present in the individual CIs, resulting in a neutralized impact on the attribute CI. Nevertheless, it is possible that our house base image was not ideal for the attributes of invitingness and likeability and lacked other relevant features commonly associated with these attributes in the context of house facades. The structure of the base image plays a crucial role in the outcome of the RC task (Brinkman, Todorov, & Doitsch, 2017), and this structural influence might be a relevant factor for attributes like invitingness and likeability. In general, participants might not have had a distinct and uniform mental representation of what constitutes an inviting or likeable house facade or there may exist multiple representations of what is seen as inviting and likeable. Since there were no significant differences in interest in art and architecture or anthropomorphism among the groups, we can conclude that these factors did not impact the differences in the outcomes of the CIs.

Comparing the quantitative number of CIs with information value, it was the largest for facelikeness, followed by invitingness, and likeability. This aligned with the observation that the attribute CI for facelikeness displayed the most distinct features. We tested for differences in sex assigned at birth and infoVal scores, however, the statistical tests did not show significant differences, which indicated that sex assigned at birth did not affect the information value of the individual CIs. We also did not observe a significant correlation between infoVal scores and neither interest in art and architecture nor anthropomorphism, which highlights that the performance in the RC task was independent of these variables.
Similarly, we did not find any significant differences between infoVal scores and participants’ living situation, in terms of their housing type. Upon evaluating the quantity of individual CIs with information value, it was observed that, on average, only a quarter of the CIs contained more than random noise. This finding supports the possibility that the base image may have been not ideal for the RC task. Since our base image is an average representation of the training dataset of the GAN, an improved base image could be generated by training on a different dataset, which consists of houses displaying other potentially relevant features. Given the limitation of 300 trials in the RC task to prevent participants from getting bored, the outcome may be more attributed to the base image rather than the experimental design.

In summary, with this work, we demonstrated the applicability of the RC task beyond the field of social psychology and succeeded in visualizing participants’ mental representation of a face on a house facade. However, further research is needed to improve the method for creating base images that effectively capture the features associated with different attributes and to examine the individual differences in the mental representations regarding architectural design.

5. Limitations

Finding a suitable base image for houses proved to be notably more challenging compared to faces, introducing complexities that may impact the effectiveness of the suggested GAN-based approach. It is possible that the base image may lack features essential for perceiving an inviting or likeable house, contributing to the observed outcomes. Moreover, the potential cultural mismatch between the base image, derived from a dataset of Californian houses, and the German participants is a noteworthy consideration. Features like a garage, prominently depicted in the base image, might not align with common architectural characteristics in Germany where garages integrated into the main part of the house are less prevalent. This cultural incongruence could have influenced participants’ mental representations and responses.

To enhance the generalizability and validity of the findings, future research could explore the same attributes with different base images, considering cultural variations in architectural norms. Further, base images could represent a broader range of building types to investigate architectural design beyond detached houses. Creating high-quality datasets containing houses and buildings from various countries is necessary for future studies to utilize the full potential of the GAN-based approach. Furthermore, investigating the importance of the base image and assessing its impact on participants would contribute to a more comprehensive understanding of the RC task in the context of house facades. Additionally, the selection of attributes for RC requires further exploration. Diversifying the attributes under examination could uncover a broader spectrum of mental representations, enriching the utility of the RC task in the assessment of architectural features.

6. Conclusion

We applied the reverse correlation method on house facades to uncover the mental representations associated with a façade, inviting, and likeable house. The resulting CI for facelikeness distinctly highlighted facial features, including eyes and a potential nose. For invitingness and likeability, no specific features of the house were accentuated in the attribute CI, however individual differences were observed in the individual CIs with a focus on the door of the house. By extending the application of reverse correlation to the area of environmental psychology and architecture, this work contributes to broadening the understanding of how individuals mentally represent and evaluate features of house facades. This expansion of the methodology’s scope opens new possibilities for future research to investigate diverse attributes and refine the approach for visualizing mental representations in architectural contexts.

Declaration of generative AI in scientific writing

During the preparation of this work the authors used ChatGPT 3.5 in order to get suggestions on how to improve the wording and phrasing of sentences and paragraphs. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Funding

Funded in part by the European Union (ERC-2022-CoG-BrainScape-101086188). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency (ERCEA). Neither the European Union nor the granting authority can be held responsible for them.

Data availability

The house base image can be requested from pohlmann@mpib-berlin.mpg.de.

CRediT authorship contribution statement

Kira Pohlmann: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. Nour Tawil: Writing – review & editing, Methodology. Timothy R. Brick: Writing – review & editing, Methodology, Conceptualization. Simone Kühn: Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was conducted at the Max Planck Dahlem Campus of Cognition (MPDCC) of the Max Planck Institute for Human Development, Berlin, Germany. During the work on her dissertation, K.P. was a pre-doctoral fellow of the International Max Planck Research School on Computational Methods in Psychiatry and Ageing Research (https://www.mps-uc-centre.mpg.de/comp2psych) of the Max Planck Institute for Human Development, Berlin, Germany. During the work on her dissertation, N.T. was a pre-doctoral fellow of the International Max Planck Research School on the Life Course (https://www.imprs-life.mpg.de/) of the Max Planck Institute for Human Development, Berlin, Germany.

References
