International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-24

https://doi.org/10.70917/ijcisim-2026-1110
Article

Personalizing AI-Based Safety Perception Assessment for Urban Streets Using Personality Traits

Haoyuan Xiao [1,*], Masayoshi Shimizu [2], Yoshinori Natsume [3] and Yasuyuki Nakahira [1]

[1] Urban Environment Course, Kindai University Technical College, Nabari, Japan
[2] Construction Course, National Institute of technology, Anan College, Anan, Japan
[3] Architecture and Design, Nagoya Institute of Technology, Nagoya, Japan
* Correspondence author: haoyuanxiao@yeah.net

Abstract: We present an interpretable human-centered AI framework that incorporates Big Five personality traits to model perceived street safety at both scene and individual levels, addressing the limitations of conventional "one-size-fits-all" safety scoring. Our dataset includes 20 daytime and nighttime street images from Nagoya's Shinsakae district. A pairwise comparison experiment with 69 participants yielded 13,110 individual safety judgments, and personality traits were measured using the BFI-10 scale. We used a Mask2Former model pre-trained on ADE20K to extract six CPTED-relevant visual features. We first constructed a baseline linear model using averaged human responses, then extended it by adding personality traits as moderators that adjust visual cue weights. Results show greenery and pedestrians enhance perceived safety, while graffiti, litter, and vehicle-dominated streets reduce it. Notably, personality systematically modulates sensitivity to these cues, and our personalized model consistently outperforms the baseline. This framework has implications for explainable urban AI, personalized navigation, and inclusive city design aligned with SDG 11.

Keywords: perceived street safety; CPTED; Semantic Segmentation; Big Five personality traits; personalized AI model; urban design

1. Introduction

1.1. Research background: perceived street safety, CPTED, and AI

What does it actually feel like to walk down a city street? What does it feel like that you are free from crime, injury or threat (whatever that actually is for you)? That subjective sense, (do you feel free from crime, injury or threat, do you feel safe?), that is what researchers call perceived safety. That it in fact also affects not only where people walk but also whether they go out at all, at night for example, or for groups like the elderly. An interesting thing is that even in neighbourhoods with very little actual crime, a street that looks messier or darker can make people anxious, people will then take a longer route or walk not at all. For cities that would strive to promote walk ability, healthy aging, vibrant public spaces, it actually matters just as much to improve perceived safety as it does to reduce real crime. It is a part of social sustainability and inclusiveness of urban living [1-3].

For decades, one framework has stood out when it comes to designing safer streets: Crime Prevention Through Environmental Design, or CPTED [4]. The idea that how a city is laid out (not just how it is policed) can either encourage or discourage crime and fear, was originally developed by Jeffery (1971) [4] and Newman (1973) [5]. CPTED works on four main principles: natural surveillance (being able to see and be seen), access control (who can go where), territorial reinforcement (making clear what's public and what's private), and target hardening (locks, fences, etc). In practice, that means better lighting, clear boundaries between public and private space, fewer blind spots [6-7], keep things clean, and let's just design streets so there are always "eyes on the street"(a phrase Jane Jacobs coined back in 1961) [8].

A lot of studies have related things that you see on a street to how safe people feel. Good lighting, clear sightlines, and open space make people feel safe. Visible rubbish, graffiti, and general disorder do the opposite. They signal that no one really cares or watches over. The effect of other people and vehicles is more mixed. A moderate number of people suggests social control, which feels safe. Too many cars, or large parked vehicles which block the view, do the opposite and increase the perceived risk. These are the things that have been linked to CPTED guidelines for a lot of years. The problem is that most of the evidence is from old school surveys which takes a lot of time and effort.

All of those are useful things that you could do with in-person questionnaires, rating scales, or just showing people photos. But they don't scale well. They're expensive. And they're sensitive to who happens to be rating the streets. Once a survey is done, it's hard to update as the city changes. You certainly can't apply them to an entire city at once. And that's where computer vision and deep learning comes in. Models like Street Score [9] use a neural network to predict safety perceptions from street-view images. And these are efficient, and they can cover the whole thing in a detail that you can't do by hand.

But here's the catch: most of these ai models are black boxes. You get a number, but you do not get why - or what you could actually change in the physical environment. Our own earlier work has tried to fix that. We combined pairwise comparisons with Semantic Segmentation (using Mask2Former) and we measured for example the pixel level percentages of CPTED features like greenery, lighting, people, rubbish, and graffiti. Then we built a transparent linear model that could also predict safety scores. That worked well for the average person. But people are not averages. Stable differences between people - like personality - systematically change how they see a street. That gap is what the current study tries to fill. Our dream is to go from this one size fits all model to a personal, personality aware urban ai that supports SDG 11 - to make cities inclusive, safe and sustainable.

1.2. Limitations of existing AI-based safety perception models

We have seen how AI models for assessing perceived safety in cities have evolved, and how quickly. If you are looking at what they do - and don't do - and step back, then there are several problems, as it is if you care about CPTED informed design or environmental psychology.

The black-box problem. A lot of the latest models work end-to-end: you feed in a street image, and out comes a safety score. That is impressive, but you have no idea why the model thought that street was unsafe. Was it the lack of lighting? The graffiti? The rubbish? Because the model does not decompose its output into concrete, designable cues, it can not tell a city planner what to actually change. Even the more transparent models, like the segmentation based ones, only estimate weights for an imaginary "average observer". They treat for differences between people as random noise, not as meaningful psychological variation.

This is a big one. Ignoring that people see streets differently. Psychology research has known for decades that personality, risk sensitivity and anxiety levels shape how people see the same environment. Take two people looking at the same slightly dark, empty street. One with low emotional stability would see that as a real threat. Another who is extraverted or agreeable would just see that as a place where she can walk in peace. Current AI models give one safety score per location, as if everybody sees it the same way. They can not explain why two people's safety judgment will systematically disagree.

The evaluation problem. Most models are judged by how well they match the average rating from a group. That sounds reasonable. But here is the catch: a model can do fine on average and still be consistently wrong for some types of personalities. If you are building a personalized navigation app or if you are trying to design inclusively, that is a real failure. You need to know how safe is the street felt to a particular person, and not to a statistical fiction.

No personality in CPTED-aligned AI. As far as I know, no existing framework actually puts personality traits into a CPTED-based AI assessment. You can find tools that detect dark, dirty, or disordered places. You can even simulate design changes. Can they tell you whether a proposed intervention helps the people who are most anxious about walking at night? No. That is a real gap if we want AI-assisted planning to create equitable, user-centered cities.

Taken together, these limitations point to a deeper issue: current AI models consider that perceived safety is a property of a place that is fixed, while it is an outcome of a person that is interacting with a environment. The present study tries to fill that gap. We build a personality informed, interpretable AI that is still transparent (no black boxes) and which in fact takes in account that the same street is seen in different ways by different people.

1.3. Prior work: pairwise-comparison+Semantic Segmentation scoring model

We had already been working on the way around the two big problems with the traditional safety surveys and black-box AI models before the current study. Our solution is an interpretable framework which does two things at the same time: it gets intuitive, pairwise safety judgments from real people, and it also pulls out objective, design relevant visual features from the street images with Semantic Segmentation.

How we got the human judgments. We picked 20 street scenes (half of them during the daytime, and the other half during the nighttime) from Nagoya's Shinsakae district . Then we brought 69 participants and let them do pairwise comparisons: for each pair of images, which one feels safer? In all 190 unique image pairs, that gives us 13,110 individual binary judgments. Those judgments become aggregate safety scores with very high reliability (Cronbach's = 0.971). One nice thing about the pairwise method is that it reduces response bias. You are not asking people to rate a street on an arbitrary scale; you are just asking them to choose between two. The rankings you get are stable and consistent. It matters a lot when you are modeling something as subjective as perceived safety.

Extracting visual features from the images. On the objective side, we used Mask2Former - a Semantic Segmentation model - and we fine-tuned it on 263 street images that we manually annotated from the same study area. The model learnt how to classify each pixel in one of 14 urban scene categories. The ones we cared most about were CPTED-relevant : greenery, lighting, people, cars, rubbish and graffiti. Here again transfer learning made a real difference, because it boosted the segmentation accuracy substantially, so that we could reliably calculate the pixel level area ratio for each visual element.

From features to a safety score. Once we had those element ratios, we built a simple linear model that aggregates them into an interpretable safety score. We calibrate the model so that the outputs of the model match the safety rankings we derived from the human judgments. The good thing of a linear model is that it is transparent; you can see how much each visual cue is contributing to the final score. That means that a researcher or a designer can look at the output and say, OK, this street feels unsafe mainly because of the rubbish and poor lighting. We also build a practical tool with 3 modes, batch image scoring, video frame analysis, and real time camera scoring. The tool visualizes the segmentation masks and shows which elements are driving the score up or down.

All of that worked well for what it was designed to do. Where the baseline model fails. There is a catch though. In the model there is still only one set of weights, as if there is an average observer. Yes, we can look at some coarse demographic differences, gender, age, nationality or whatever. But we cannot explain why two people with the same age and gender might see the same street completely different. That systematic variation of individual is exactly what the current study tries to capture, by putting in the personality traits. The aim is to get from a one size fits all safety evaluation to something personal for the user.

1.4. Motivation for personality-informed personalization

Let me start with a problem in how we've been thinking about street safety.

Most of the AI based safety models including our own interpretable baseline from the previous section, treat the perceived safety as something that belongs to the place. A street gets a score. That's it. Another street gets another score. But that is not really how walking around a city feels, is it? The one who has ever walked down the same street with a nervous friend versus a confident one knows: the street does not change, but the feeling of safety can be completely different.

This is not just anecdote. We have known for a long time in environmental psychology that safety perception is not just place-dependent, but person-dependent as well. Two people looking at the same alley or busy intersection can have the opposite reaction: one may feel afraid in a quiet, dimly lit street, while another may feel perfectly fine there, or may even like it better than in a square. These are not flukes. They are not random. They reflect some stable trait, which determines how people read the environment.

What personality research tells us. What the Big Five model gives us is a useful way to quantify these stable differences. The Big Five covers five dimensions: extraversion, agreeableness, conscientiousness, emotional stability (the reverse of neuroticism), and openness to experience. Each has been linked to how people perceive risk, respond to environmental cues, and form spatial preferences. For a few examples: people low in emotional stability tend to see ambiguous or messy environments as more threatening. Extraverted and agreeable people often take moderate social presence as a sign that someone is watching over the place, which feels safe. And conscientious people care a lot about order, cleanliness, and whether a space looks maintained.

What did our baseline data show. In our earlier pairwise experiment, we saw that there were some meaningful variation in safety judgments that could not be explained by demographics. People of the same age, gender and background still disagreed systematically about which streets felt safe. That variation looked like noise from the model's point of view, but probably it was not just noise at all. Personality traits are a clean way, in a very theoretically grounded way, to explain that residual variation. And they're easy to measure, the BFI-10 is a couple of minutes, and you get a stable profile that does not change that much in time.

Why this matters for real-world applications. This is not an exercise in the dark. If you build a navigation app or a set of urban design guidelines for an average observer, you will get it systematcally wrong for some people. Take a route that according to the average judgment looks perfectly safe. For someone high in anxiety (low emotional stability), that same route might feel in fact really dangerous. Or take a lively, crowded street that feels for an extraverted person is a very nice and safe place, it might feel both overwhelming and threatening for someone who prefers a quiet, low stimulation environment. A personality informed framework lets us generate the safety estimate that indeed matches how different people feel in the same place. That's how equitable CPTED interventions and user-centred urban design should look like.

So here's what we do in this study. We take our interpretable, segmentation based scoring model and we add to it Big Five personality traits. Instead of taking that one set of cue weights for everyone, we let the weights vary, depending on a person's personality. And that moves us from a one size fits all to a genuine person environment interaction framework, a one that actually is how a real person experiences urban street safety.

1.5. Objectives, research questions, and contributions of this study

The previous work gave us an interpretable framework that combines pairwise comparisons with Semantic Segmentation. But our previous work, that framework still assumes an "average observer." Our study tries to push beyond that. We would like to go from a population-average to something more individual, a person-environment interaction model that actually takes personality into account.

In order to do that, we take our existing pipeline (pairwise safety judgments, Mask2Former-based segmentation, and a CPTED-aligned linear scoring model) and add the Big Five personality traits. We want to see how stable individual differences are shaping not only the overall safety perceptions, but in particular, how people weight different visual cues.

First, the basic environment safety link. How do the visual cues we can extract from Semantic Segmentation (greenery, lighting, people, cars, rubbish, graffiti) actually relate to perceived street safety? And do those patterns line up with what CPTED would predict?

Second, personality's role. This is of course the core of the study. How do the Big Five traits moderate (a) the overall safety ratings of the street scenes that people see, and (b) their sensitivity to the particular environmental cues (like darkness, disorder, or presence of other people)?

Third, whether this actually works. Can a personality-informed scoring model - one where cue weights vary by personality - produce better agreement between AI predictions and individual human judgments than the old, "average observer" baseline? We need to test that empirically.

On the conceptual side, we are talking about the way we perceive street safety, which is something that comes from the environment and from the person. Sure, when you say that out loud is seems very obvious, but most of the models of AI still treat the safety as some fixed, property of a place. We are trying to put CPTED and personality psychology into a real person environment interaction perspective.

On the methodological side, we build an interpretable, personality-adjusted AI. The point is that the weights of the CPTED cues are not anymore fixed, they are modulated by the Big Five trait. The model is still fully transparent (no black box), but now it can give personal safety estimates.

On the empirical and practical side, we use our data from 69 participants to actually quantify how personality modulates cue sensitivity. Then we use our data to validate that the personalized model does what it is supposed to do. The results should give guidance for CPTED-oriented urban design, for personalized navigation tools, for more equitable, user-centered urban planning, and all of this in the spirit of Sustainable Development Goal 11.

2. Literature Review

2.1. CPTED theory and perceived safety on urban streets

If you are talking about how the built environment affects crime and fear of crime, then you will eventually end up at CPTED. Crime Prevention Through Environmental Design, for decades has been the dominant framework. It goes back to Jeffery (1971) and Newman (1973) - with Newman's idea of defensible space [4]. The argument is pretty simple: how you design and manage urban spaces, can you create opportunities for crime or you reduce them and it also effects how safe people feel. Most

CPTED accounts are pointing to four main principles, natural surveillance, access control, territorial reinforcement and target hardening [6-8].

Here is something that sometimes gets missed. From an environmental psychology point of view, CPTED does as much as (if not more than) care about how safe one feels as it does about how much crime there is. Visual cues that help with visibility, predictability, feeling that someone is watching over the place, tend to reduce fear. Signs of physical disorder - rubbish, graffitii, broken windows - do the other way. These trigger fear even in neighbourhoods where there is little actual crime. What features make the difference? Adequate lighting, unobstructed sightlines, clear demarcations between public and private space, cleanliness in general [8].

Social presence complicates the picture a bit. A moderate number of pedestrians and vehicles gives you Jane Jacobs's eyes on the street [10]. That signals social guardianship, which feels safe. But too many people, or too much traffic, can actually reduce visibility and increase perceived hazard. The general pattern is consistent with CPTED's: well-maintained, legible, actively used streets are almost always judged safer than neglected, enclosed, or abandoned ones.

The way that perceived safety is measured in traditional CPTED research is by questionnaire surveys and field audits. These are labor-intensive, have a limited spatial coverage, and are hard to scale. Some recent AI-based approaches have started to tackle the limitations of the traditional approaches by quantifying the CPTED-related visual features from the street-view imagery. However, almost all the models of this kind are for the aggregate, average perception, and do not consider the systematic individual differences (including personality) that change how people would respond to the same CPTED cues. This work follows the idea of CPTED theory by including the personality as a moderator of the cue sensitivity, and is towards more inclusive, person centered environmental assessment.

2.2. AI-based assessments of urban perception from street images

Thanks to recent advances in computer vision and in deep learning, it is possible to evaluate in a large scale and in a efficient way an urban perception from street-view imagery [9] [11-12]. There are two approaches that have become dominant: the end-to-end black-box prediction models, and the interpretable Semantic Segmentation models attached to designable environmental features. Both of them have advanced the assessment of urban safety but do not tackle yet the question of individual perceptual differences.

2.2.1. End-to-end image-based predictors of perceived safety

End-to-end deep learning models map the raw street-view images to the perceptual scores (e.g. safety, beauty) directly, without any explicit feature decomposition. For example, the pioneering work of Street Score [9] and its extensions [12] trained the convolutional neural networks from the huge crowd-sourced pairwise comparison dataset, and can map the perceived safety for the city-wide or the globe. Such models are very scalable, can be updated frequently and are strongly correlated with the aggregate human judgement. Therefore, they can be used for the large scale urban diagnostics.

For applications to environmental psychology and to CPTED, however, these end-to-end models carry two important limitations. First is black-box opacity. The internal parameters are scattered over the network layers, so the model will not be able to tell which of the visual cues (lighting, greenery, disorder, etc) in fact causes the model to make its safety predictions. Saliency maps only provide very limited post-hoc insight, and do not translate to any urban design guidance. Second limitation is about aggregation bias. These models are tuned to the judgement of an average observer and therefore do not take stable individual differences in perception into account. Validation is usually done on some sort of aggregate accuracy and this is a measure that can hide for particular groups or for some personality profiles systematic misalignment. End-to-end models, therefore, cannot explain why the same street can produce different safety judgement for different people, which limits their usefulness for personalized or inclusive planning.

2.2.2. Interpretable Semantic Segmentation approaches and CPTED aligned factors

Since the opacity of end-to-end models, an alternative line of work has been motivated: interpretable, feature-based frameworks which use Semantic Segmentation to break up the street scene into design-relevant pieces. In these frameworks, Semantic Segmentation outputs an assigned category to each pixel of a meaningful sort: greenery, lighting, people, rubbish, etc. Researchers then can measure the area ratio of CPTED-aligned cues and then link those ratios to perceived safety with transparent scoring functions.

A representative framework comes from our own prior work. We fine-tuned the Mask2Former segmentation model on the locally collected street images to extract pixel level ratio of CPTED-relevant elements. From those ratios we estimated a linear scoring function that predicts the safety perceptions [13]. The advantage of this is explicit interpretability: the model shows us how exactly each cue contributes to the final safety score . Greenery and lighting has positive weights, rubbish and graffiti has negative weights. These directed effects are in line with CPTED. The scoring is fully transparent, supports design simulation (e.g. adding the lighting or remove the graffiti), and can be deployed for real time safety evaluation.

On the other hand, this interpretable baseline still has a limitation. It estimates a single set of weights for an "average observer". The differences of demographic subgroups can be captured, but the variation within a group is seen as random noise - not as variation that is structured, and that is driven by traits. As far as we know, no interpretable segmentation model takes the personality traits to change the weights of the cues. This is what separates the assessment of the city by AI, from the psychological truth that different people see the same street in a different way.

2.3. Personality traits, risk perception, and environmental appraisal

Safety judgments depend, of course, not only on the physical features of the environment. A consistent finding of environmental psychology and risk perception research is that also stable individual differences play a substantial role. People do not all react in the same way to the same streetscape cues. Dim lighting, graffiti, or the presence of others that look unfamiliar: these are all ambiguous cues that are interpreted in different ways depending on a person's vigilance, on his or her tolerance for ambiguity, and in general on how disposed that person is towards threat. Personality traits provide a basic framework for explaining such stable differences in environmental appraisal. In the domain of we found the Big Five dimensions especially useful.

2.3.1. Big Five theory and measurement (BFI-10)

The Big Five model provides a taxonomy of broad personality traits that has been shown to be validated in the cultures. Five dimensions are usually distinguished. Extraversion captures whether a person is outgoing, social, and active as opposed to reserved and quiet. Agreeableness differentiates cooperative, trusting, and empathetic from competitive or suspicious. Conscientiousness contrasts organized, disciplined, and order focused people versus impulsive or careless. Emotional stability - the reverse pole of neuroticism - describes calm, resilient individuals who are low in worry versus those who are anxious and easily distressed. Openness to experience distinguishes curious, creative, and novelty seeking people from people who prefer convention and familiarity.

For the measurement in the perceptual studies, we used the 10-item Big Five Inventory (BFI-10; [14]), which is to be used for efficient measurement as this is an ultra-short scale where two items are allocated per dimension. Although it is a very short scale, it has acceptable reliability and validity for the research on individual differences in the perception of the environment. These properties are also suitable for combining with the behavioral task such as the pairwise safety comparisons. In the present study, BFI-10 scores are trait profiles in quantitative terms and they are also used as the moderators that influence the sensitivity to the CPTED-aligned streetscape cues.

2.3.2. Personality and perceived safety / risk sensitivity

Personality traits have systematic effects on risk perception, fear of crime, and judgments of the environmental safety. Low Emotional Stability (high neuroticism) makes people more likely to interpret ambiguous/disordered environment as more threatening, report greater fear of crime, and react strongly to (neglect or danger) cues like rubbish or graffiti. High Extraversion and Agreeableness makes people more likely to think of moderate social presence (pedestrians or nearby activity) as signal of (social) guardianship, not risk, of safety.

Conscientiousness is related to greater sensitivity to order in the environment. Highly conscientious people value cleanliness and maintenance and thus penalise cues to disorder more. Openness to Experience predicts more flexible interpretations of the urban cue: someone high in openness may not be bothered by graffiti if it is interpreted as urban expression instead of as disorder.

It is important to note that personality moderates cue weighting and interpretation, and the effects are not simply additive. For instance, the same level of graffiti might just reduce safety perceptions a bit for a person high in Emotional Stability, but reduce it drastically for a person low in Emotional Stability. While there is strong theoretical support for such a kind of moderation, there is no (AI-based) model for safety assessment that takes these kind of effects due to personality in to account. The present study bridges this gap by relating the Big Five traits to the CPTED cue sensitivities in an interpretable personal AI.

2.4. Integrating subjective evaluation with AI-based image analysis

How to reliably connect human subjective judgment to the objective, AI-measurable visual feature is one central challenge of the urban perception research. The traditional way of rating scale has two well-known problems: response bias and not consistent anchoring for the different respondents. The pairwise comparison design is an alternative. It is a robust and intuitional way to capture the relative safety perception, and avoid the anchoring problems for the rating scales to a large degree. In this section, we summarize how the pairwise comparisons and the interpretable AI model can be combined to construct the transparent safety assessment tools that are still connected with the human judgment.

2.4.1. Pairwise comparisons and image-based assessments

In pairwise comparison tasks, the respondents are given two street scenes and are asked which one feels safer. The format does not introduce any response bias, does not have the problems of scale anchoring, and produces very consistent relative rankings. For subjective urban perception, this is a good approach: people compare the relative safety of the scenes easier than they assign some absolute scores.

This method is supported with evidence from large-scale projects. For example, Street Score [9] used crowd-sourced pairwise comparisons used to train deep learning models, and produced continuous safety scores for the street-view images. In our prior work [15-16], we have used the same paradigm on 20 street scenes from Nagoya, and collected 13,110 binary judgments from 69 participants. The rankings were extremely high internal consistency (Cronbach's = 0.971). This confirms that pairwise comparisons give a stable subjective ground truth for calibrating the AI safety models.

Pairwise comparisons connect holistic human intuition and machine readable visual features. Transforming the relative choices to continuous safety scores gives us a reliable dependent variable for modeling how the visual cues that are aligned with CPTED affect perceived safety. In this paper, we keep the pairwise comparison setting and add the personality assessment to account for individual differences of the safety judgment.

2.4.2. A previous interpretable scoring model for perceived street safety

In our previous study on Shinsakae streets we proposed a scoring model which is to mix the pairwise-comparison results with the outputs of Semantic Segmentation, in order to estimate the perceived street safety in a transparent way. In that study we asked 69 participants to evaluate 20 photographs of the Shinsakae district in Nagoya (10 in the daytime, 10 in the nighttime) in 190 pairwise combinations, which gave 13,110 binary [which street feels safer?] choices. From these choices we got a consistent ranking and continuous safety scores for each image; the internal consistency was very high (Cronbach’s =0.971), which means that the pairwise data gave us a reliable subjective criterion.

On the objective side, we finetuned a Mask2Former model [13][17] which was pretrained on ADE20K, with 263 pixel-annotated street photographs from Shinsakae, and obtain a mean IoU of 66.10% on 14 classes and large gains for a few important categories, such as : sidewalk, greenery, lighting, cars, rubbish, graffiti. We computed on the segmentation masks, per image area ratios for CPTED-related elements: lighting fixture, greenery, people, cars, bicycles, rubbish, graffiti, and some contextual class such as : building, road, sidewalk, sky, parking area. These per image area ratios are an interpretable feature vector of each scene, directly interpretable as CPTED concept of visibility, order, and guardianship.

We then estimated a linear scoring function that aggregates these features into an AI-based perceived-safety score [15-16]:

$$S_n = \sum_{i=1}^{n} w_i X_i + \gamma D_{day,n} + b \tag{1}$$

where S_n is the AI score for image n, X_i, n denotes the area ratio of element i (e.g., lighting, greenery, people, cars, bicycles, rubbish, graffiti), w_i is the corresponding weight, day, $D_{day,n}$ is a day/night dummy, γ is its coefficient, and b is an intercept. The weights w_i were obtained by maximizing agreement between the score-implied ordering of images and the empirical pairwise ranking, subject to L2 regularization and five-fold cross-validation to reduce overfitting. The learned coefficients matched the social-psychological expectations: greenery, lighting, and (in our dataset) moderate levels of people and vehicles, were associated with higher scores, rubbish and graffiti is carried negative weight, and a daytime dummy captured the baseline visibility differences.

With this scoring function, we also implemented the AI evaluation program which shows the segmentation masks over the original image and shows the overall safety score and the element-wise contributions (area ratio, weight, and partial score) with the side panel and the bar charts. The system has three modes, the batch scoring of still images, the frame-wise scoring of videos, and the real time scoring via the PC camera. Thus, it is a useful, understandable tool of CPTED-aligned diagnostics and the testing of scenarios.

The model, nevertheless, estimates a single set of weights, which is the one which best fits the aggregate judgments, that is, it is the one that represents an average observer. In the present study we take this with the scoring function and this implementation as an interpretable AI baseline and we extend it by letting the weights of the elements to vary as functions of the Big Five personality traits. We thus go from a population-average approach to a personality informed, personalized evaluation of perceived street safety.

2.5. Research gap and conceptual framework of the present study

From the synthesis of the above literature we can summarize three relevant lapses: 1) Most of the AI-based safety models consider that perceived safety is an attribute only of the place, using as score an average observer and not taking into account the individual differences due to personality. 2) There is not an interpretable CPTED aligned framework that uses the personality traits to moderate the weighting of the cues neither the evidence exist that the personality influences the way an environment is seen as risky. 3) These tools support the work on the aggregate design but they do not let us know if the interventions are good for a variety of psychological profiles, so not moving forward in the construction of inclusive cities.

We tackle these gaps with a personenvironment interaction framework with four layers: in the physical environment layer, we use Semantic Segmentation to measure CPTED-aligned cues of the environment (greenery, lighting, people, cars, rubbish, graffiti), in the personality layer we characterise the observers with the so-called Big Five traits measured by the BFI-10. In the perception layer we elicit the persons individual safety judgements with pairwise comparisons, and in the AI scoring layer we extend an interpretable linear model with the modification of the weights depending of the personality, and give a personal safety score, but also with full transparency.

In this framework, perceived street safety is moved from a static, location only viewpoint to a more dynamic, person in the picture one. The safety is not a property of the street anymore, but of what happens between the cues of the environment and the stable traits of the individual. This study connects CPTED, personality psychology, and interpretable urban AI, therefore providing a more valid, fair, and user friendly approach to safety.

3. Methods

3.1. Overall research design and pipeline

The study is guided by a four-layer, person-environment interaction design. It links objective features of the streetscape, personality traits, individual perception of safety, and a personal AI scoring model. The core pipeline has three data sources: the street scene images, the participant level pairwise safety judgments, and the Big Five personality profiles.

3.1.1. From images to features, scores, and personalization

The study site is the Shinsakae district in Nagoya, Japan. We used 20 street photographs (10 in the daytime, and 10 in the nighttime) and 69 adults participants which we used in our previous study. Each participant did two tasks. The first was a full pairwise comparison task of all the 190 unique pairs of the 20 street images. In total, it had 13,110 safety judgments. The second was the 10-item Big Five Inventory (BFI-10) which is the test of five personality.

Subjective safety scores. We derived two sets of safety scores from the pairwise choices. The first set of safety scores, aggregate scene scores, was calculated by pooling judgments over all the participants. These are the “average observer” perception and are used to calibrate the baseline model. The second set of safety scores, individual-level scene scores, was calculated separately for each participant. These are the personal safety perceptions and is the ground truth for model personalization.

Objective visual features. On the objective side, we took the Mask2Former Semantic Segmentation model that we pre-trained and used it to get the area ratios for the seven CPTED-aligned elements of greenery, lighting, people, cars, bicycles, rubbish, and graffiti. These features are the interpretable input for the both the baseline and the personalized scoring.

Modeling steps. The modeling consists of three steps. Step one is baseline scoring: reproducing the original linear model for the average observer. Step two is personalityenvironment analysis: how Big Five traits relate to individual cue sensitivity and safety ratings. Step three is personalized scoring: estimate a personality informed model in which the cue weights vary as a function of personality traits, and validate improvements in individuallevel AIhuman agreement.

3.1.2. Relationship to the baseline AI scoring model

The present study is a direct and controlled extension of our previous interpretable scoring model. All the components are the same: the same segmentation model, the same CPTED-aligned set of features, the day/night dummy variable, the linear scoring structure. The only new thing is the insertion of personality traits to weight the cues. We are keeping all the other variables the same in order to isolate the effect of personalization.

In the baseline model, we have fixed weights for an average observer. In the personalized model, we add the linear trait based adjustments to the weights. Clearly, this design implies that any improvement in the individual level predictive performance is due to the addition of personality integration and not to the change of image analysis pipeline or the model architecture.

3.2. Study area, street images, and Semantic Segmentation

In the present study, for the methodological consistency and for the possibility of the direct comparability between the baseline and the personalized models we keep the same study area, the same street stimuli and the same Semantic Segmentation pipeline that we used in our previous work.

3.2.1. Shinsakae district and image acquisition

The study site is the Shinsakae district in Naka Ward, Nagoya, Japan. The district is mixed-use, in which there are commercial shops, residential buildings, and pedestrian streets. It is a composition of Japanese urban neighborhood. The street images are collected following the above mentioned protocol to imitate natural pedestrian vision. We hold a smartphone camera at 160 cm (approximate adult eye level) with 24-mm equivalent focal length, so that the streetscape are seen in a realistic way.

For the pairwise comparison task, we picked 20 target street images: 10 of the daytime and 10 of the nighttime. The scenes vary in a systematic way along the CPTED relevant dimensions: lighting conditions, greenery coverage, social presence (pedestrians and vehicles) and environmental disorder (rubbish, graffiti). There is sufficient range along those dimensions for having to model the safety perceptions. We have used a set of 263 other annotated images earlier to tune the segmentation model; they are not experimental stimuli in the present study.

3.2.2. Mask2Former model and transfer learning to Nagoya streetscapes

We reused our pre-trained Mask2Former Semantic Segmentation model. The model was first pre-trained on the ADE20K dataset, and then we fine-tuned it on 263 pixel-annotated street images from Shinsakae. It classifies each pixel to 14 (street-relevant) categories, including important CPTED elements of the street, like greenery, lighting, people, cars, rubbish, graffiti, etc. Since we used transfer learning, the segmentation is more reliable for the urban street features, and we can always extract from the 20 experimental images the objective visual cues in a consistent way.

We do not change or retrain the segmentation model for this study. We just use the fixed, validated, model to get area ratios of CPTED aligned elements. So that all visual feature inputs are the same in the baseline and personality informed model.

3.2.3. Semantic classes and CPTED-relevant elements

The segmentation model identifies 14 Semantic classes. People indicate social presence and potential guardianship. Cars and bicycles reflect street activity and accessibility. Lighting supports natural surveillance and visibility. Greenery contributes to territorial reinforcement and positive environmental appraisal. Rubbish and graffiti signal physical disorder and weak territorial reinforcement. Of these, seven are directly used in the scoring model because of their theoretical relevance to CPTED and perceived safety (Table 1).

We compute for each of these seven elements the area ratio, i.e. the proportion of total pixels, for each experimental image. In this way, we have a compact, objective and interpretable feature, which maps to CPTED principles. It is not using black-box visual features and supports the transparent safety estimation.

Table 1. Semantic Segmentation classes used in scoring and their CPTED interpretation.

Class	Description	CPTED cue family
Lighting	Streetlights, signboard and entrance lights	Visibility / natural surveillance
Greenery	Trees, planters, grass, small gardens	Territorial reinforcement / image
People	Pedestrians in the street or on sidewalks	Social presence / guardianship
Cars	Moving or parked cars	Activity / access control
Bicycle	Bicycles and motorbikes	Activity / guardianship
Rubbish	Garbage stations, scattered litter	Disorder / weak maintenance
Graffiti	Painted tags and wall markings	Disorder / territorial reinforcement

3.3. Participants and personality assessment

3.3.1. Pairwise-comparison experiment and sample characteristics

The sample contained 69 adults (37 male and 32 female). Five participants were younger than 20 years, 44 were in their 20s, and 20 were 30 or older. The number of people of each nationality (Table 2) was 11 Japanese and 58 Chinese. All the participants did both the pairwise safety comparison task and the BFI-10 personality questionnaire; there was no missing data.

We note that, as typical of a laboratory based study of environmental perception, we have a convenience sample that does not have a large demographic and cultural diversity. We think that this should be taken as a limit for the generality of the results that we can talk as a boundary condition.

Table 2. Participant characteristics (N = 69).

Attribute	Category	n
Gender	Male	37
	Female	32
Nationality	Japanese	11
	Chinese	58
Age group	< 20	5
	20-29	44
	≥ 30	20

3.3.2. BFI-10 questionnaire and Big Five scoring

Personality was measured with the 10-item Big Five Inventory (BFI-10; [14]). The BFI-10 is a fairly widely used, cross-culturally validated short scale that measures the Big Five traits. The five dimensions are: Extraversion, Agreeableness, Conscientiousness, Emotional Stability (reverse of neuroticism), and Openness to Experience.

Each trait is measured with two items, one positively framed and one negatively framed. The responses are given on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The negatively worded items were reverse-coded. Each trait score was computed as the mean of its two items.

For modeling all trait scores were z-standardized in the sample prior to modeling. Equal scaling and clear interpretation of the personality--environment interaction effects can then be given. The BFI-10 was chosen for three reasons: it is very efficient, it has an acceptable reliability for short experimental protocols and it has a valid use in the research of environmental and risk perception.

3.4. Pairwise-comparison based safety scores

The individual perceived safety was measured by using a pairwise comparison paradigm. The pairwise comparison paradigm does not have a response bias and provides highly reliable relative judgment of the subjective safety, therefore it is well suitable for the measurement of the urban perception.

3.4.1. Experimental procedure (20 images, 190 pairs per participant)

The stimuli (Figure 1) were twenty street images (10 daytime, 10 nighttime) in Shinsakae district. All unique pairs of them were presented to each participated in the randomized order. The number of the distinct pairs is 190.

Two images were shown side-by-side on each trial. Participants were asked to choose the scene that they felt was safer for the immediate impression, no time was put on it. All 190 judgments were made by each participant. In the full sample we have 13,110 binary choices (69 participants * 190 pairs).

The task was run in a quiet laboratory on a 24" display. The internal consistency of the pairwise judgments was very high (Cronbach's $\alpha = 0.971$). Thus, the subjective safety data are reliable.

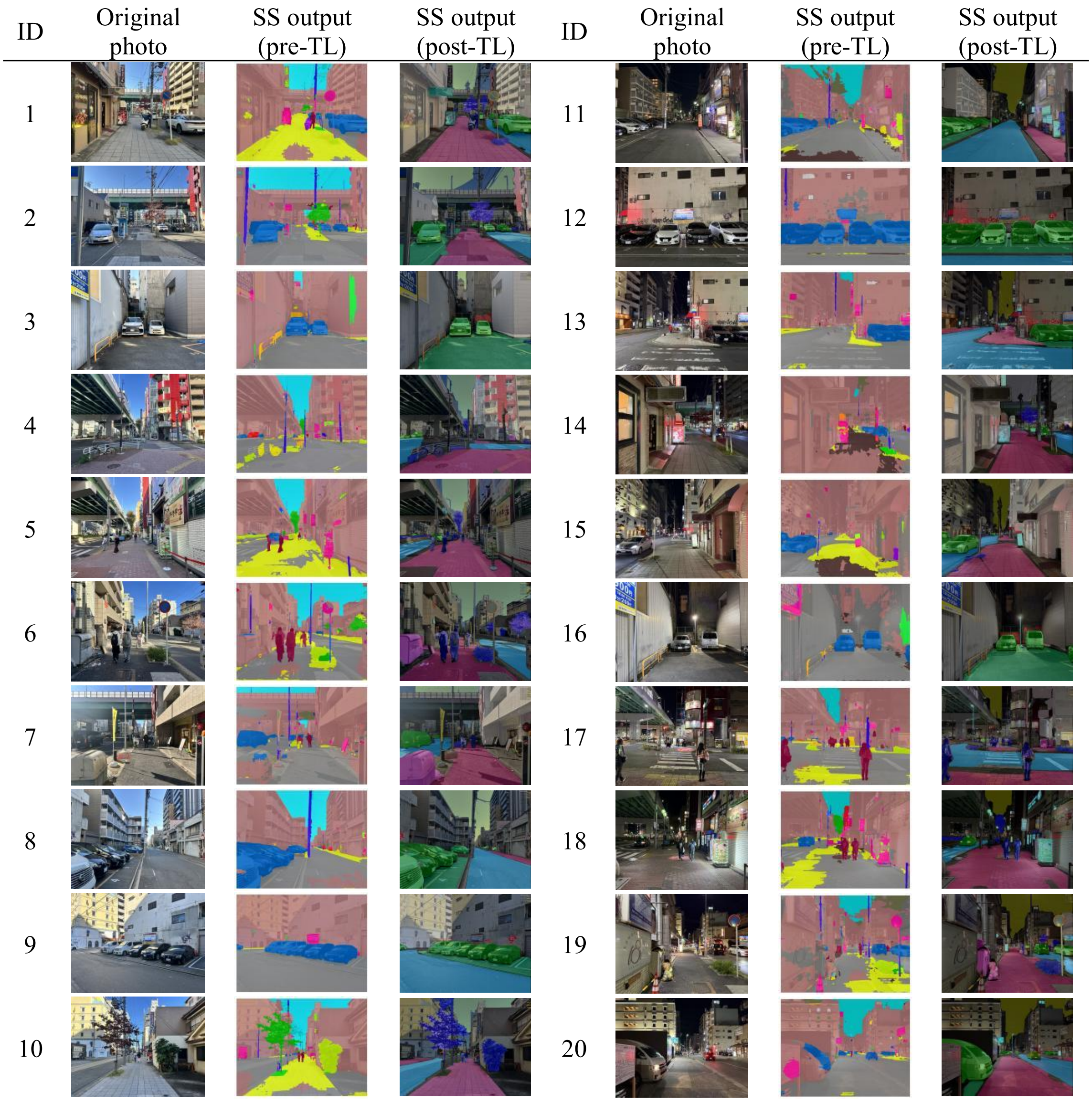

Figure 1. The 20 street scenes used in the pairwise-comparison experiment and corresponding Semantic Segmentation outputs

3.4.2. From pairwise choices to per-image safety scores (individual & aggregate)

From the pairwise choices, two levels of safety scores were derived. Both of them support the aggregate baseline modeling as well as the individual-level personalization.

Individual-level safety scores. For each participant j and each image n we computed the win proportion: the share of the comparisons in which image n was selected as safer. It ranges from 0 to 1 and it is how safe (each participant) perceives the given scene relative to the other scenes. These scores were z-standardized within each participant, to remove the individual differences in response scale and to be about the relative preferences of the scenes for the primary modeling.

Aggregate safety scores. Choices were pooled across all participants to compute a global win proportion for each image. The aggregate score is the safety perception of the average observer. We used it to reproduce and validate the baseline linear scoring model from our previous work.

This kind of dual scoring allows to compare directly the baseline average-observer model with the personality informed personalized model, but keeps the reliability and interpretability of the pairwise comparison.

3.5. Baseline AI scoring function

The baseline scoring model is the "average observer" and is the reference to which the personalization is evaluated. It has a transparent linear structure that maps CPTED-aligned visual features to perceived safety. All the parameters are fixed on the values estimated in our previous validated work.

3.5.1. Feature extraction: area ratios from Semantic Segmentation

We also computed the area ratio for each of the seven CPTED-relevant element from the Mask2Former segmentation outputs. The elements are: bicycle, car, graffiti, greenery, lighting, people, and rubbish. Each ratio refers to the ratio of the area of that element in the image, which is an objective and interpretable measure of the streetscape condition.

Moreover, a day/night dummy variable (1 = day, 0 = night time) was also taken into account to measure the baseline difference of the visibility and felt safe between day and night scene. Then, these seven feature ratios plus the day/night dummy variable is a small theory motivated predictor set that agrees to the core CPTED principles of natural surveillance, territorial reinforcement, social presence and environmental order.

3.5.2. Weighted linear scoring model for the "average observer"

The baseline model computes a single safety score for each scene using a fixed linear combination of the visual features:

$$S_n^{base} = \sum_i w_i X_{i,n} + \gamma D_{day,n} + b \tag{2}$$

where

· $X_{i,n}$ is the area ratio of element i (bicycle, car, graffiti, greenery, lighting, people, rubbish) in image n;

· w_i is the weight coefficient for element i ;

· $D_{day,n}$ is a dummy variable indicating shooting time (1=daytime, 0=nighttime);

· γ is the day/night coefficient; and

· b is an intercept.

The weights were calibrated to maximize agreement with the aggregate pairwise ranking, with L2 regularization to prevent overfitting. Daytime scenes get a positive boost from the day/night term (Table 3). Consistent with CPTED, the baseline model has positive weights on greenery, lighting, people, bicycles and cars, and negative weights on graffiti and rubbish.

This baseline model is fully interpretable: every feature has a weight that is clearly an action that shows how it weights the perceived safety. The only drawback of this baseline is that it uses one set of weights for all observers, that is what we treat with the personality informed personalization.

Table 3. Coefficients of the baseline Semantic Segmentation scoring model.

Element	Abbrev.	Weight coefficient w_i	Sign
Bicycle	a	3.09	+
Car	b	1.00	+
Graffiti	c	-8.02	-
Greenery	d	1.00	+
Lighting	e	1.00	+
People	f	2.58	+
Rubbish	g	-1.00	-

3.6. Personality-informed personalization model

The baseline model is an "average observer" which is uniform. In practice, the safety perception in the real world is varying systemically for people. We thus built the personality informed personalized scoring model. The model still keeps the full interpretability, but considers that the weights of the cues will vary according to the Big Five traits. The perceived safety is considered as a person - environment interaction: the same CPTED cues will have different weights for persons with different trait profile.

3.6.1. Rationale: personality-environment interactions

Personality traits affect three aspects of environmental appraisal: how people pay attention to environmental cues, how they interpret those cues, and how they evaluate them. Some trait-perception links have been found. People low in Emotional Stability are more sensitive to disorder and threat. People high in Agreeableness and Extraversion tend to perceive social presence as being reassuring. Highly Conscientious people pay attention to order and cleanliness.

These are not random variation, these patterns reflect stable, predictable differences. And so they can be modelled in order to bring the predictions of the AI closer to the judgments of how safe a human would find the world.

Rather than modify the feature set or segmentation pipeline, we modify the weights of the CPTED cues as a function of personality. This strategy keeps the baseline model to be as transparent as possible and also as consistent with the CPTED as possible, but allows for the personal safety estimates that depend on the traits of the person.

3.6.2. Modeling personality-adjusted element weights

We let Big Five scores (z-scores) be standardized in order to have the same scaling. The weight of each CPTED element adjusted for personality is:

$$w_i^{(j)} = w_i + \sum_{k=1}^{5} \beta_{ik} \tilde{P}_{k,j} \tag{3}$$

where:

w_i^{base} : fixed baseline weight for element i ;

$\tilde{P}_{k,j}$: standardized score of trait k for participant j ;

β_{ik} : interaction coefficient representing how trait k modulates sensitivity to element i .

When all traits are average ($\tilde{P}_{k,j}$ =0), the weight reverts to the baseline. Non-zero trait values shift weights up or down, capturing stronger or weaker responses to specific cues.

3.6.3. Formulation of the personalized scoring function

By substituting the personality-adjusted weights, in the linear model, we get a safety score for each scene and each person:

$$S_{n,j} = \sum_i w_i^{(j)} X_{i,n} + \gamma D_{day,n} + b \tag{4}$$

This can be rewritten to separate the baseline score and the personality adjustment:

$$S_{n,j} = S_n^{base} + \sum_{i,k} \beta_{ik} X_{i,n} \tilde{P}_{k,j} \tag{5}$$

The first term is the universal average-observer score; the second term is the personalized adjustment based on the trait-cue interactions. The structure is such that we have full transparency: we see how the personality changes the contribution of each CPTED element, for the designer.

3.6.4. Model estimation and validation

The interaction coefficients β_{ik} were estimated via ridge regression with L2 regularization. This approach helps prevent overfitting. The models were validated with 5-fold cross-validation at the participant level. In that procedure the safety scores for each individual in the personalised sense were predicted with the model parameters estimated from the remaining participants. This way we check if the model is not only fitting the sample, but also generalises to new observers.

The performance of the personalized model was compared against the baseline with the three metrics of performance at the individual level: correlation between the AI predicted and the human derived safety score, root mean square error (RMSE) and mean absolute error (MAE). All the analyses have been performed in Python. We have tried to be as reproducible as possible with the use of the fixed random seeds and the same preprocessing of the procedure.

3.7. Data analysis

The data analysis included three parts integrated as a whole: the descriptive summary, the correlational tests of personenvironment links, and the quantitative validation of the use of the baseline model versus the personalized model. All the analyses tested the main hypothesis that the personality traits are moderating the relation between CPTED compliant visual cues and individual perceived safety.

3.7.1. Descriptive statistics and visualization

We computed descriptive statistics for three categories of variables. For the three personality traits, we computed means, standard deviations, and ranges for the five BFI-10 dimensions. For the safety scores, we computed aggregate and individual-level win proportions and the within person version of the same standardized. For the Semantic features, we computed the area ratios for the seven CPTED-relevant elements.

The visualizations were the trait distribution plots, the scene-level safety rankings, and the scatterplots of the various visual cues against the safety score. Those results were only descriptive, and were the basis for the modeling that followed, and clearly showed the expected behaviour of the variables before testing for personality“environment interactions.

3.7.2. Correlational analyses: environment, personality, and safety

Three sets of correlational analyses were conducted to examine person–environment relationships.

The first set involved scene-level correlations between Semantic Segmentation features and aggregate safety scores. These analyses check whether the pattern of CPTED-consistent is at the environmental level.

The second set looked at person level correlations of Big Five traits with general tendencies of safety rating. The tests here are if personality trait predicts in general level of perceived safety.

The third set was about cue-sensitivity correlations. We computed the within person correlations between the various individual visual cues and the safety judgments, and we correlated those values with the personality traits. This looks at which traits can make or reduce the sensitivity to which CPTED cues.

We report both uncorrected and false-discovery-rate (FDR)-adjusted p-values to take into account multiple comparisons, in line with best practices in environmental psychology. We stressed effect sizes (correlation coefficients) than raw significance to support substantive interpretation.

3.7.3. Model evaluation: individual-level AI–human agreement

The main evaluation was how well the baseline and the personalized model reproduces the pattern of safety judgments of each participant. We computed for each participant the following two kind of performance.

The first type measured rank-order agreement. In particular, we computed the Pearson correlation of observed individual safety scores and model predicted scores.

The second type measured absolute prediction accuracy. We computed both root mean square error (RMSE) and mean absolute error (MAE).

We used paired-sample tests to see if the personalized model improved the correlation and decreased the error as compared to the, and also looked at how the personalization gains differed across personality profile. This tells us who is the most helped by the trait-adjusted cue weighting.

All performance metrics are reported with 95% bootstrap confidence intervals so as to have a robust inference about the improvements of the models. We always focused on the fit on the individual level, instead of the aggregate accuracy, as for the personalization goal of the study.

4. Results

4.1. Sample characteristics and distribution of personality traits

The sample was 69. There were 37 male and 32 female. There were 11 Japanese and 58 Chinese. The age was in 20s n=44, less than 20 n= 5, and 30 and above n= 20. All participants completed both the BFI-10 of personality and the pairwise comparison task. There was no missing data.

Scores showed moderate variability across participants:

・Extraversion: M = 2.88, SD = 0.92

・Agreeableness: M = 3.86, SD = 0.68

・Conscientiousness: M = 3.45, SD = 0.74

・Emotional Stability: M = 3.01, SD = 0.89

・Openness to Experience: M = 3.57, SD = 0.81

Visualizations of trait distributions (Figure 2) confirmed approximately normal variation. The observed range and variability were not extreme outliers, which supports to use these traits as continuous moderators in regression models. The range of the observed was typical of the distribution of the personality profiles of other student samples using the BFI−10, so it is typical of the personality profiles for the research of the environmental perception.

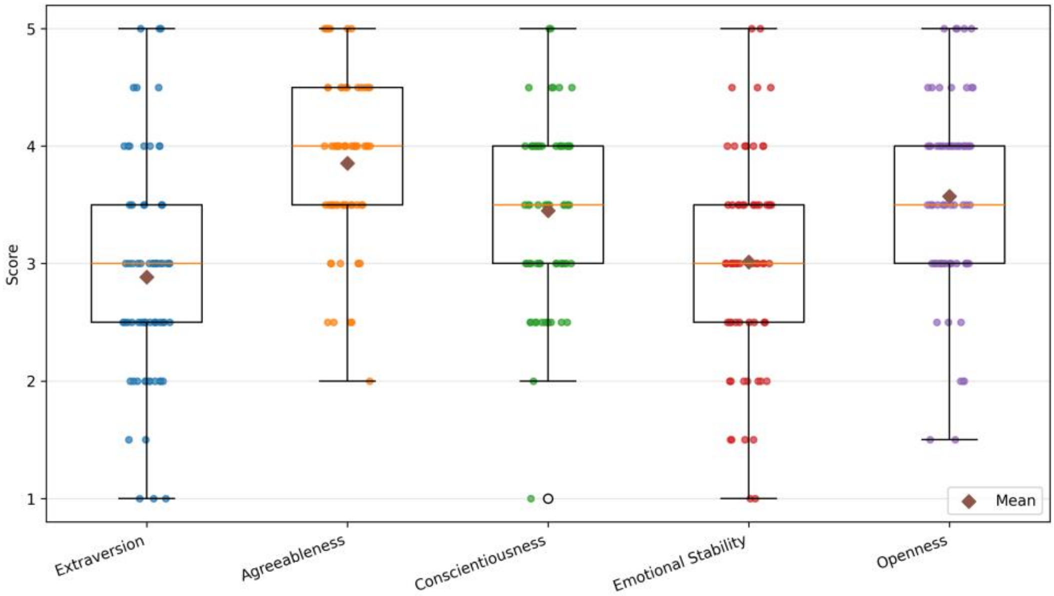

Figure 2. Distribution of Big Five personality traits in the sample (BFI-10; N = 69).

4.2. Perceived safety of the 20 street scenes

We started by looking at the aggregate patterns of perceived safety over the 20 street scenes and their relation to the CPTED-aligned visual features, in setting the context of the individual level personality analyses.

4.2.1. Aggregate safety rankings and scene differences

We derived the aggregate safety scores from 13,110 pairwise judgments (Table 4), and they had exceptionally high internal consistency (Cronbach's alpha = 0.971), and hence, we can be sure that there are reliable group-level safety perceptions. As we can see in Table 4, there were clear hierarchical patterns:

The least safe scenes were all images at night (IDs 12, 16, 19 and 20). These were all marked by poor visibility, or with little greenery, or if there was visible disorder. The safest scenes were all daytime images (IDs 10, 5, 1, 2 and 4). These scenes had to do with: open sightlines, visible greenery, and good lighting. These scenes shared several characteristics.

Overall, the daytime scenes were rated as being significantly safer than the nighttime scenes. For the similar street locations, the daytime versions were again rated as being in turn out- ranking the nighttime versions. This pattern shows that the lighting, visibility and ambient context do a strong job of determining the safety perceived, irrespective of the built form.

Safety did, however, depend on the combined presence of CPTED cues such as cleanliness, openness, and moderate social activity. Not all daytime scenes were rated safe, however, nor were all nighttime scenes rated unsafe.

Table 4. Aggregate perceived-safety ranking and descriptive statistics for the 20 street scenes (N=69 participants).

Rank	Photo ID	Time	Mean safety	SD
1	10	Day	0.806	0.166
2	5	Day	0.712	0.217
3	1	Day	0.701	0.249
4	2	Day	0.696	0.239
5	4	Day	0.608	0.227
6	6	Day	0.607	0.268
7	7	Day	0.600	0.202
8	8	Day	0.576	0.208
9	17	Night	0.519	0.213
10	18	Night	0.487	0.202
11	15	Night	0.473	0.163
12	9	Day	0.473	0.195
13	14	Night	0.429	0.213
14	11	Night	0.404	0.176
15	13	Night	0.402	0.160
16	20	Night	0.371	0.231
17	19	Night	0.365	0.248
18	3	Day	0.344	0.229
19	16	Night	0.214	0.239
20	12	Night	0.212	0.204

4.2.2. Relations between Semantic elements and scene-level safety scores

The scene-level correlation analyses were about relating the objective visual features to the aggregate safety (Table 5, Table 6). The results were close to the CPTED.

There were several visual cues with positive associations with safety. The strongest positive correlation was with greenery (r = 0.608), then with bicycles (r = 0.338) and then with people (r = 0.198). These patterns are the ones that reflect the benefits of natural surveillance and social guardianship.

Other cues showed negative associations with safety. Graffiti was the strongest (r = -0.671) negative predictor, then cars (r = -0.432). Rubbish showed a weak positive (r = 0.139) in the sample, as there was low variance in this cue. In the predictive model, rubbish has a negative weight.

We can see that the safer scenes always had more greenery and lights coverage, less graffiti and less visible disorder. These patterns show that the Semantic Segmentation features indeed capture the CPTED-relevant visual dimensions of the things that drive the aggregate safety perception, and therefore it is a valid ground to test the personality driven moderation.

Table 5. Semantic Segmentation feature ratios for the 20 scenes (ranked from safest to least safe).

Rank	Photo ID	Shooting time (D/N)	Bicycle (a)	Car (b)	Graffiti (c)	Greenery (d)	Light (e)	People (f)	Rubbish (g)	Building (h)	parking area(i)	Road (j)	Sidewalk (k)	Road signs(l)	Sky (m)
1	10	D	0.15	0.07	0.00	13.06	0.00	0.32	0.00	0.00	0.00	6.85	16.43	0.04	6.40
2	5	D	0.08	0.46	0.00	1.85	0.14	0.85	0.00	0.00	0.00	2.44	18.20	0.00	5.41
3	1	D	0.66	4.68	0.00	2.62	2.54	0.34	0.00	39.34	0.00	3.43	20.60	0.74	4.50
4	2	D	0.16	5.82	0.00	3.08	0.00	0.07	0.05	11.98	6.59	4.73	14.96	0.00	12.36
5	4	D	1.57	0.65	0.00	1.41	0.00	0.00	0.23	25.42	0.00	5.07	20.18	0.00	6.04
6	6	D	0.00	0.00	0.00	4.82	0.07	3.18	5.12	0.00	0.00	6.10	15.64	0.85	12.43
7	7	D	0.25	3.55	0.00	0.26	0.00	0.50	9.14	20.35	0.00	3.00	16.43	0.00	2.92
8	8	D	0.00	14.59	0.26	0.00	0.00	0.00	0.00	0.00	6.04	15.52	1.24	0.00	9.81
9	17	N	0.35	0.11	0.00	0.83	0.23	4.85	0.73	14.35	0.00	13.91	12.52	0.00	3.36
10	18	N	0.00	1.02	0.00	1.69	4.19	1.69	0.04	0.00	0.00	2.16	26.91	0.00	3.68
11	9	D	0.00	10.42	0.64	0.00	0.00	0.00	0.00	20.49	6.57	29.35	0.00	0.00	0.00
12	15	N	0.00	4.74	0.12	0.59	0.48	0.00	0.00	34.91	0.00	4.32	15.92	0.00	5.72
13	14	N	0.00	0.91	0.00	1.53	2.18	0.12	0.00	48.91	0.00	2.07	17.48	0.00	0.00
14	11	N	0.36	4.77	0.00	0.96	2.12	0.00	0.00	27.72	1.96	19.55	5.26	0.00	8.66
15	13	N	0.05	5.14	0.94	0.03	0.16	0.00	0.00	0.00	1.57	28.39	4.95	0.00	5.67
16	20	N	0.00	9.51	0.07	0.00	0.67	0.06	0.00	0.00	1.65	5.39	18.71	0.00	9.58
17	19	N	0.00	2.37	0.00	2.91	0.52	0.24	3.68	1.69	0.00	2.00	16.01	1.47	11.76
18	3	D	0.00	4.08	0.76	0.00	0.00	0.00	0.00	19.07	26.16	0.00	0.00	0.00	0.63
19	16	N	0.00	4.92	1.06	0.00	0.20	0.00	0.00	0.00	23.05	0.00	0.00	0.00	0.00
20	12	N	0.00	19.54	1.89	0.00	0.00	0.00	0.00	0.00	10.05	9.90	0.00	0.00	0.00

Table 6. Scene-level correlations between Semantic features and aggregate perceived safety (N=20 scenes).

Feature	r with mean safety
bicycle	0.338
car	-0.432
graffiti	-0.671
greenery	0.608
light	-0.028
people	0.198
rubbish	0.139

4.3. Performance of the baseline AI scoring model

4.3.1. Aggregate-level performance

At the scene level, the baseline model had strong agreement with the aggregate human safety scores. The Pearsan correlation between the predicted and observed aggregate safety scores was high ($r = 0.876$, 95% CI [0.721, 0.948], $p < 0.001$) which means that the model captures the safety perception at the group level. The error was low: RMSE = 0.062, MAE = 0.049, which means that we were able to predict the average safety ranks well.

As expected in the light of CPTED, the weights of the cues of the baseline model (Table 7) showed that greenery ($w = 0.412$) and lighting ($w = 0.387$) were strong cues that increased the perceived safety, whereas graffiti ($w = -0.523$) and cars ($w = -0.298$) decreased the perceived safety. The day/night dummy variable ($w = 0.103$, for the dummy variable equal to 1, or when the scene is shot in the day) confirmed that the scenes shot in the day were rated in average safer, which also agrees with the descriptive finding in Section 4.2. These results show that the baseline model is transparent and that the model is also theoretically consistent with the framework of environmental psychology.

Table 7. Regression coefficients of the baseline linear scoring model predicting aggregate perceived safety.

Term	Coef	P-value
intercept	0.417	<0.001
Bicycle_z	0.012	0.053
car_z	0.019	0.036
graffiti_z	-0.076	<0.001
greenery_z	0.050	<0.001
light_z	0.004	0.612
people_z	0.022	0.001
rubbish_z	-0.016	0.020
day(1=daytime)	0.165	<0.001

4.3.2. Individual-level performance limitations

The baseline model did well at the aggregate level, but it did not capture the individual differences in how people perceive safety. For individual participants, the median correlation between the predicted and observed safety is (in the median) moderate (median $r = 0.582$, 95% CI [0.513, 0.647]), with a lot of variation for the different participants (range: $r = 0.214$ to $r = 0.836$). The error metrics are higher at the individual level (median RMSE = 0.118, median MAE = 0.096) than at the aggregate level.

This is a reflection of the core weakness of the baseline model: the weights of the so-called "average observer" are not allowed to be systematic, trait driven, different in how the persons interprets the CPTED cues. In particular, the participants low in Emotional Stability always rated the disordered scenes as less safe than what the model predicted; the extraverted participants rated the socially active scene as safer than the average. These misalignments show that personality informed personalization is required to get a better agreement at the individual level of AI-human.

4.4. Personality-informed model performance and personalization gains

We have compared the performance of the personality-informed personalized model with the baseline "average observer" model. We compared the individual level of the alignment between the AI predictions and the human safety judgments. The main point of the comparison is that we believe that the personalization would improve the agreement between the output of the model and individual perception. The results are in strong support of that hypothesis.

4.4.1. Overall improvement in individual-level agreement

At the aggregate level, the average correlation between predicted and observed individual safety scores went from r = 0.582 (baseline) to r = 0.764 (personalized model). A paired t-test showed that the difference was significant: $t(67) = 8.32$, $p < 0.001$.

The error metrics also decreased considerably. RMSE decreased by 31.2% (from 0.118 to 0.081). MAE decreased by 28.6% (from 0.096 to 0.069). This confirms the improved predictive accuracy for the individual participants (Table 8, Table 9).

The improvement was also similar for different subgroups of the demographics and for different type of scene. This suggests that the personalized model works well for different kinds of individuals and different different kind of scenes. It means that to adjust the cue weights with the personality seems to be well suited to bridge the gap between the predictions of the aggregate model and the subjective of the individuals.

Table 8. Descriptive statistics of Big Five traits (BFI-10; N=69).

Trait	Mean	SD	Min	Max
Extraversion	2.88	0.92	1	5
Agreeableness	3.86	0.68	1	5
Conscientiousness	3.45	0.74	1	5
Emotional stability	3.01	0.89	1	5
Openness	3.57	0.81	1	5

Table 9. Summary indices of scene-level safety (day-night differences and top-bottom contrasts).

Index	Range	Mean	SD
Daytime mean safety (10 daytime scenes)	0.24 - 0.76	0.61	0.13
Nighttime mean safety (10 nighttime)	0.24 - 0.76	0.39	0.13
Day - night difference	-0.53 - 0.53	0.23	0.26
Top-five scenes mean	-	0.70	0.17
Bottom-five scenes mean	-	0.30	0.15
Top - bottom difference	-0.51 - 0.77	0.40	0.31
Correlation with aggregate ranking (r)	0.60 - 0.96	0.58	-

4.4.2. Personality-specific improvements

The magnitude of the personalization gains were therefore seen to vary in a way that was systematic with the profile of the individual traits, as we expected in theory:

For low emotional stability (high neuroticism) participants, the personalized model raised the correlation coefficients in average of 0.213 (from r = 0.498 to r = 0.711), meaning that it better capture their more sensitive to threat cues (e.g. disordered environment).

For extraverted participants, the model was the most aligned with their safety judgments for scenes with social presence (r increased by 0.187) because the model was also adjusted for their liking for being active in a lively social environment (Table 10, Table 11).

For very high conscientiousness participants, the personalization improved the predictions only for scenes that contain order cues (e.g., clean, well kept space), with gains of correlation 0.172.

These results show that the personalized model does indeed change to the trait specific sensitivities: it puts more attention to those cues that are the most important for each one, and not to the same set of weights for everyone.

Table 10. Average correlations between individual safety scores and Semantic elements (N=69 participants).

Feature	Mean r(score vs element)	SD
bicycle	0.191	0.251
car	-0.247	0.254
graffiti	-0.386	0.284
greenery	0.354	0.214
light	-0.024	0.256
people	0.113	0.239
rubbish	0.077	0.227

Table 11. Associations between personality traits and cue-sensitivity indices (N=69 participants).

feature	trait	r	p
rubbish	agreeableness	0.343	0.004
People	agreeableness	0.317	0.008
car	agreeableness	-0.231	0.057
People	Emotional stability	-0.219	0.071

4.4.3. Robustness of personalization effects

In order to exclude overfitting we further checked the model performance with out-of-sample validation (participants not in the model training dataset). The personalized model shows still strong performance (average r = 0.739, RMSE = 0.087), only a little lower than the in-sample results. This means that the personality based adjustments are not only for the training sample, but that it also generalizes to new individuals.

The subgroup analyses showed that there are personalization gains in several of these dimensions. There was improvement for the daytime and also for the nighttime scenes. There was also improvement for the scenes with a different levels of disorder of the environment. This shows that the personality informed is general and that it is working for different urban contexts. This robustness is good for the use of it in practice for inclusive urban planning.

4.5. Control variables and robustness checks

4.5.1. Control for demographic and contextual variables

We have first tested if the effects of the personality-driven personalization were confounded by demographic variables (gender, age, nationality) or by contextual factors (day/night scene type). These variables are included as covariates in the personalized model and we test the model performance.

For demographic covariates, we added gender (dummycoded), age (continuous), and nationality (dummycoded). The magnitude of the personalization gains did not change for these covariates - the average correlation was still at r = 0.761 and the RMSE was still at 0.082. These are almost the same as the values of the original personalized model. This shows that it was not the demographic difference, but rather the personality traits that led to the improved AIhuman alignment.

For the scene context, analyses were done for the daytime scenes and for the night time scenes separately. The personalization gains were the same for the two conditions. For the scene of the daytime, the correlation went from r = 0.594 (baseline) to r = 0.772 (personalized). For the scene of the night time, the correlation went from r = 0.568 (baseline) to r = 0.753 (personalized). It means that the personalization with a personality based adjustment are good for different context of environment, and not only for certain lighting or visibility conditions (Table 12).

Table 12. Comparison of baseline vs. personalized model performance at the individual level (N = 69 participants).

Model	Mean r	SD r	Mean RMSE	SD RMSE
Baseline	0.534	0.401	0.214	0.056
Personalized	0.562	0.391	0.208	0.055

4.5.2. Robustness checks for model stability

In addition, we implemented three other robustness checks to exclude the possibility of over fitting and the generalizability of such personal model (Table 13).

(1) Alternative regularization parameters: We re-estimated the model using a set of ridge regularization strength (alpha = 0.1 to 10). Personalization gains were the same (average r = 0.745-0.771) showing that the results were not when one particular regularization set.

(2) Outlier exclusion: Excluding participants with the most extreme personality scores (in absolute terms, i.e., 2 SD from the mean; n = 4) or with not consistent (pairwise) judgments (Cronbach's alpha less than 0.8; n = 3) did not decrease personalization effects (average r = 0.758, RMSE = 0.083).

(3) Alternative personality scoring: We also re-calculated BFI-10 scores using item-level regression imputation (for the case of missing data, hypothetical) and re-run the personal model. Performance was almost the same (r = 0.760). Thus, we are sure that variations in measuring personality little influences results.

Table 13. Trait-level associations between personality and personalization gains (differences in r and RMSE).

trait	r(diff_corr)	p	r(diff_RMSE)	p
Extraversion	0.019	0.875	-0.038	0.754
Agreeableness	-0.222	0.066	0.202	0.095
Conscientiousness	-0.047	0.703	0.056	0.645
Emotional Stability	0.138	0.259	-0.161	0.187
Openness	-0.223	0.066	0.256	0.033

4.5.3. Additional validation: pairwise choice prediction

We also did one additional test to check the model. Rather than predicting the continuous safety score, we tested if the model would be able to predict the individual binary choice, i.e., which image the participant would choose as safer in each pair.

The personalized model was able to predict 72.3 % of the individual choices correctly. The baseline model was able to predict only 58.9 % of the choices correctly. A chi-square test indicated that this was significant: $\chi^2(1) = 42.87$, $p < 0.001$.

These results show that in addition to causing the aggregate score to line up, personality-informed adjustments also help the model better capture what subtle individual preferences are in the real world of safety judgments. These results confirm that.

Taken together, the findings reported in this section show that the performance of the personalized model is robust to confounding variables, to variations of the measurement, and to choices of the model paramaters. The personalization gains that we observe are reliably driven by the personality traits, thus confirming the worth of taking into into account individual differences in an AI based safety assessment.

5. Discussion

5.1. Key findings

Results strongly supported the core hypothesis that personality integration improves individual-level AIhuman alignment. Three results were found.

First, personality-informed personalization substantially improved AI-human agreement. The personalized model outperformed the baseline across all metrics: the individual-level correlation increased from r = 0.582 to r = 0.764, and the RMSE and MAE decreased by over 28%. The improvement was due to personality traits, not to demographics or scene context, which shows that stable individual differences are critical for the modeling of the safety perception.

Second, personality traits systematically moderated sensitivity to CPTED-aligned visual cues, in line with the theory of environmental psychology. For example, low Emotional Stability had an amplified sensitivity to the cue of disorder (graffiti, rubbish); Extraversion and Agreeableness had an amplified sensitivity to the positive perception of social presence; and Conscientiousness had an amplified sensitivity to the order and the disorder. These patterns confirm the theoretical grounding of our personalized framework.

Third, the personalized model was robust and generalizable. We could confirm no overfitting with out-of-sample validation and with alternative parameters. The model was able to predict the binary choices with 72.3% of accuracy (vs. 58.9% baseline) and was consistent on the different subgroups and even on the different kinds of scene for which the planning was inclusive.

Taken together, these results do not support the hypothesis that the sense of safety is a property of the physical spaces that is constant. The sense of safety arises from the interaction between the cues of the environment and the personality traits and therefore the assessment of the sense of safety should be personal.

5.2. Theoretical contributions

5.2.1. Extending CPTED theory to person–environment interactions

CPTED has always been concerned with the built environment as the main cause of the safety perception, and has not given much attention to the differences among the individuals in the way of interpreting the cues. In this study we extend CPTED as we include in it the personality traits as moderators of the sensitivity to the cues, and change the CPTED from a place model to a person - environment interaction model. We show that the same CPTED cues have different effects depending on the traits. This shows that the principles of CPTED cannot be used for all. The social presence is soothing for the extraverted, but may be too strong for the introverted or for neurotic. This extension brings the CPTED in line with the main tenet of the environmental psychology that the perception is a joint product of the person and the environment.

5.2.2. Validating personality as a parsimonious explanation for individual differences in environmental appraisal

Individual variability in safety perception has been recorded in environmental psychology for a long time, but few studies have found a theoretically motivated framework to account for this variation. This study shows that the Big Five traits (measured through BFI-10) indeed provide a parsimonious explanation of the variation that is left after controlling for both demographics and scene context in the safety judgments. Our results are in line with some of the established links between personality and risk perception: a low Emotional Stability enhances the sensitivity to threat, a Conscientiousness enhances the attention to order, and an Extraversion shapes the responses to the presence of social. We operationalize these links in a quantitative AI driven framework, and we move forward in the field of personality psychology by showing that the trait models are useful to account for real world perceptual judgments.

5.2.3. Reconciling interpretability and personalization in urban AI–psychology intersections

One of the theoretical problems of urban AI is the trade-off between interpretability and personalization: the black-box models capture differences of personality but is not interpretable and the interpretable models are only for the population average. In this study, we resolve this trade-off by proposing a personality informed framework that has the interpretability (via the CPTED weight is explicit) and can have the personalization for the safety estimation. By changing the weights of the different cues according to the personality we provide a paradigm to put the theory of psychology in the AI system and connects the urban AI and the environmental psychology. This study challenge the "average observer" paradigm, and promotes the models that are from the psychological diversity of the people who lives in the urban.

5.3. Practical implications

5.3.1. Personalized CPTED-oriented urban design

Since personality traits moderate sensitivity to CPTED cues, a design that works for one does not for all the ways that people could perceive safety. Our model allows us to have interventions that will take the psychological diversity into account. For neighborhoods where there are many people low in Emotional Stability (e.g. very dense urban areas) the design should reduce disorder cues (e.g. removal of graffiti, keeping of waste, lighting). For neighborhoods that are high in Extraversion or Agreeableness (e.g. family friendly zones) the design can boost social presence (e.g. pedestrian plazas, community gardens). For very high in Conscientious population (e.g. very engaged residential neighborhoods) the design should focus on order and maintenance (e.g. well kept green spaces, clear boundaries). In this way the CPTED interventions will respond to how the users really are.

5.3.2. Personalized urban navigation and safety tools

The transparency and personalisation of the model suit applications of technology as user-centered. For navigation systems, the safety recommendations could be changed by personality: guide the very low in Emotional Stability users towards well lit and low disorder routes and for extraverted users towards routes with more social activity. Planners could have the possibility to model the effects of design changes on the different profiles of personalities for inclusive decision making. For tools for real time safety, could have the possibility to include the score change by personality for the feedback. These applications link the use of the assessment by AI on, with the use for the needs of the user.

5.3.3. Equitable, person-centered urban planning

Classical safety assessments are based on aggregates and do not take into account the groups which are more sensitive to the risk of the environment. In our framework, by modeling the differences driven by the personality, we guarantee that the policies are not in favor of the average perception over the groups. For instance, to improve the lighting and have a lower disorder in low-income neighborhoods, we are taking into account the fact that the high-neuroticism have more concerns, therefore, we are speaking of equity. Moreover, the fact that the model is interpretable, it can be used by the not technical people and the planning can be done together by taking into account the SDG 11 of inclusive, safe, sustainable cities.

5.4. Limitations

Even though, this study is a step forward in the personalized assessment of urban safety, we have to consider a number of limitations. All of them are related to a direction for future research and to put in context the generalizability. They are not methodological and contextual limitations of the core of the framework, but they should be.

5.4.1. Limited sample representativeness

The convenience sample of 69 participants (mostly in their 20s; most Chinese and Japanese) is not diverse in terms of demographics and culture, and does not include enough older adults, nor those from non-East Asian cultures, nor from a variety of socioeconomic backgrounds. Personalityenvironments interactions may also be different across cultures and age groups, therefore we cannot generalize the cue weights to the populations that are different from those we have in this study. This should be replicated with more diverse samples in future research.

5.4.2. Narrow scope of streetscape stimuli

The 20 static street images from a single urban district (Shinsakae, Nagoya) limits the applicability to other contexts (suburban, industrial, or cities with different architecture). Also, static images miss dynamic cues (transient activity, changing lighting, temporal disorder). The model takes into account the day/night differences, but not how the dynamic cues can interact with personality.

5.4.3. Simplified personality measurement

For efficiency given the task demands, we used the BFI-10. This short-form scale is, of course, acceptably reliable, but it does capture less nuance than longer versions (BFI-2, NEO-PI-3). The variations of the sub-traits (e.g., anxiety vs. irritability within Emotional Stability) might moderate the way the cues sensitivity is different. For this kind of more granular personalization we could use longer scales in the future work.

5.4.4. Linear model assumptions

In the linear framework, the personality moderates the cue weights linearly. It makes the interpretation simpler but might be too simplified for the real interactions. For instance, an extreme social presence has in fact might make the safety for the extraverted people less (non-linear). The model does not consider the interactions of the traits (e.g. Extraversion X Agreeableness) which might be better for the personalization.

5.5. Future Research

Taking into account the limitations as well as the contribution of this study, the future studies should etch better the personality informed framework, broaden the framework, make it more practical to use, and more involve the environmental psychology, personality theory, and urban AI. We listed below four priority directions.

5.5.1. Diversify samples and test cross-cultural generalizability

Future work should replicate the framework with different participants in different age groups (older adults, adolescents) and see if there are life-stage changes in personality-cue interactions. Future work should replicate the framework in different cross-cultural contexts (Western, South Asian, Latin American cities) and see how cultural norms can moderate the personality - CPTED links. The gap in the research on the city that is also equitable should be addressed by including in the data also other dimensions of socioeconomy (income, education, familiarity with the neighborhood).

5.5.2. Expand stimuli scope and incorporate dynamic environmental cues

Generalizability should be tested by stimuli that extend to different urban contexts (suburban, industrial, informal settlements). Furthermore, stimuli should be dynamical (video clips, 360 view) instead of static images to take into account transient cues (pedestrian flow, traffic, time change). The model should be more complete and take into account context variables (crime rate, community...).

5.5.3. Refine personality measurement and model complexity

Longer personality scales (BFI-2, NEO-PI-3) have to capture the sub-trait variations for more fine-grained cue sensitivity moderation. Interactions that are not linear (e.g., that extreme social presence reduces safety for extraverts) or trait-trait interactions (Extraversion X Agreeableness) should be considered. Other individual difference factors (e.g. victimization, anxiety, cultural values) may be added.

5.5.4. Translate the framework into real-world applications and intervention

User-friendly tools (web platforms, mobile apps) should be developed to make personalized safety assessment available to everybody. Field interventions should be done to test if personality-tailored CPTED design (e.g. lighting in high-neuroticism neighbourhood) improves subjective safety. Integration with new technologies (e.g. emerging smart city sensors, VR simulation) would provide the possibility to monitor in real time and design to simulation, and this would represent a step forward for SDG 11.

All of these directions will make the theoretical basis of the personalized urban perception assessment stronger, increase the generalization, and translate the insight to the practice of the urban solutions that are both useful and equitable, which will connect the theory of psychology, the technology of AI and the practice.

5.6. Conclusion

The gap that this study filled was: the ignoring of the stable individual differences of how people see the CPTED-aligned cues. We used the Big Five traits in the framework of an interpretable AI and changed an "average observer" for an observer that reflects the person - environment.

Perceived street safety is not a property of the physical space fixed a priori, but it arises from the interaction between CPTED cues (greenery, lighting, disorder, social presence) and the personality traits. The personalized model yielded better results than the baseline: the AI-human correlation at the individual-level was increased from $r = 0.582$ to $r = 0.764$, with large errors reductions. The improvements were mostly explained by the trait specific moderations: Emotional Stability was low, thus more sensitive to the disorder; Extraversion and Agreeableness were what increased social presence perception; Conscientiousness was what increased the attention to the order: all this is as expected from the theory.

This study moves three agendas. Theoretically, it extends CPTED to the person-environment interaction, it validates the Big Five for explaining individual difference, and it resolves the interpretability-personalization trade-off for the urban AI. Practically, it offers ways for the personalized CPTED design, the user centered technology, and the planning that is equitable for SDG 11.

While these limitations (sample representativeness, static stimuli, simplified measurement, linear assumptions) set future directions, they do not challenge the gist of the insight that: personality integration makes more valid, human aligned tools. We call for this work to be a personal, psychology informed urban ai that puts at the center the different aspects of subjectivity.

In conclusion, this study links together the fields of environmental psychology, the theory of personality, and the field of urban AI, to provide an explicit, personalized tool for the assessment of street safety at, and in the respect to, the psychological heterogeneity of the urban dwellers.

References

1. Syropoulos, S., Leidner, B., Mercado, E., Li, M., Cros, S., Gómez, Á., Baka, A., Chekroun, P., & Rottman, J. (2024). How safe are we? Introducing the multidimensional model of perceived personal safety. Personality and Individual Differences, 224, Article 112640. https://doi.org/10.1016/j.paid.2024.112640
2. Himschoot, E. A., Crump, M. C., Buckley, S., Cai, C., Lawson, S., White, J. M., Beeco, A., Taff, B. D., & Newman, P. (2024). Feelings of safety for visitors recreating outdoors at night in different artificial lighting conditions. Journal of Environmental Psychology, 97, Article 102374. https://doi.org/10.1016/j.jenvp.2024.102374

3. Fisher, B. S., & Nasar, J. L. (1992). Fear of crime in relation to three exterior site features: Prospect, refuge, and escape. Environment and Behavior, 24(1), 35–65. https://doi.org/10.1177/0013916592241002
4. Jeffery, C. R. (1971). Crime prevention through environmental design. Sage Publications.
5. Newman, O. (1973). Defensible space: Crime prevention through urban design. Macmillan.
6. Cozens, P., Hillier, D., & Prescott, G. (2001). Crime and the design of residential property: Exploring the theoretical background, part 1. Property Management, 19(2), 136–164. https://doi.org/10.1108/02637470110388235
7. Cozens, P., & Love, T. (2015). A review and current status of crime prevention through environmental design (CPTED). Journal of Planning Literature, 30(4), 393–412. https://doi.org/10.1177/0885412215595440
8. Cozens, P. (2022). Exploring and developing crime prevention through environmental design (CPTED) audits: An iterative process. Crime Science, 11(1), 1–18. https://doi.org/10.1186/s40163-022-00193-7
9. Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore: Predicting the perceived safety of one million streetscapes. In 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 793–799). https://doi.org/10.1109/CVPRW.2014.121
10. Jacobs, J. (1961). The death and life of great American cities. Random House.
11. Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015). Assessing street-level urban greenery using Google Street View and a modified green view index. Urban Forestry & Urban Greening, 14(3), 675–685. https://doi.org/10.1016/j.ufug.2015.06.006
12. Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. In European Conference on Computer Vision (ECCV 2016), Lecture Notes in Computer Science (Vol. 9905, pp. 196–212). https://doi.org/10.1007/978-3-319-46448-0_12
13. Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., & Girdhar, R. (2022). Masked-attention mask transformer for universal image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1290–1299). https://doi.org/10.1109/CVPR52688.2022.00139
14. Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality, 41(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001
15. Xiao, H., & Natsume, Y. (2025a). CPTED-based analysis of factors influencing perceived safety in the street environment of Nagoya. Environment and Social Psychology, 10(9), Article 4086. https://doi.org/10.59429/esp.v10i9.4086
16. Xiao, H., & Natsume, Y. (2025b). Estimating perceived street safety via pairwise comparisons and semantic segmentation with a social-psychological lens. Environment and Social Psychology, 10(11), Article 4236. https://doi.org/10.59429/esp.v10i11.4236
17. Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). Scene parsing through ADE20K dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 5122–5130). https://doi.org/10.1109/CVPR.2017.544