Understanding Pedestrian Safety Compliance in Smart Traffic Systems: A Multimodal Data Approach

Yuting Zhang[1], Jiayu Chen[1], Tao Cui[2], Yichen Liu[3], and Mingxuan Liang[4]

[1] Department of Civil Engineering, Tsinghua University, Beijing 100084, China.
[2] Mercedes-Benz Group China Ltd., Beijing 100176, China.
[3] Commercial Aircraft Corporation of China Ltd, Beijing, 102211, China.
[4] Department of Built Environment, School of Design and Engineering, National University of Singapore. SDE2, 2 Architecture Dr, Singapore 117565

yuting-z22@mails.tsinghua.edu.cn, jiayuchen@mail.tsinghua.edu.cn, tao.t.cui@mercedes-benz.com, liuyichen3@comac.cc, mingxuanliang@u.nus.edu

Abstract -

This study investigates the safety compliance of pedestrian within smart traffic systems using a multimodal data approach. By integrating behavioral records, questionnaires, eye-tracking metrics, and electroencephalography signals, the study examines the cognitive mechanisms underlying pedestrian decision-making during street crossing. The research utilizes structural equation modeling to assess how environmental factors, individual safety propensities, and perceptual strategies influence pedestrian safety performance. The results highlight the mediating role of cognitive constructs such as risk awareness, demonstrating that pedestrians' safety behaviors are shaped by complex cognitive evaluations rather than direct environmental stimuli. These findings offer insights into how traffic safety can be improved through environmental design and targeted training to enhance pedestrians' cognitive skills and decision-making processes.

Keywords -

Pedestrian safety; multimodal data; structural equation model; risk awareness; eye-tracking; electroencephalography.

1 Introduction

Safety is widely recognized as a fundamental component of smart city resilience, with traffic safety occupying a central position due to the essential role of transportation systems in urban life. Nevertheless, the safety performance of road traffic systems remains a major concern, as traffic accidents continue to cause substantial losses to human life and social well-being. Among all road users, pedestrians are particularly vulnerable due to their lack of physical protection and the high uncertainty of their behaviors. To mitigate pedestrian-related risks, many jurisdictions have adopted punitive enforcement mechanisms, such as penalties for pedestrian or driver violations, as a means of preventing pedestrian accidents. However, these existing enforcement mechanisms are largely reactive, relying on punitive mechanisms rather than addressing the underlying behavioral factors that lead pedestrians to make unsafe decisions. In fact, despite having a clear understanding of traffic rules, pedestrians often commit violations in diverse situations, which consequently exposes them to potential hazards [1]. This difficulty highlights the challenge of developing effective human-factor interventions without a clear understanding of pedestrians' motivations and the factors influencing their behavior. Given the limitations of externally imposed constraints, many existing studies have shifted their focus toward the underlying causes of pedestrians' unsafe actions. With the rapid advancement of intelligent transportation systems and autonomous driving technologies, the ability to accurately predict pedestrian behavior has become a critical requirement for ensuring system safety and reliability. In real-world traffic environments, pedestrians represent one of the most uncertain and dynamic elements, whose decisions are shaped not only by external conditions but also by internal cognitive processes. Future smart traffic systems are expected to incorporate predictive capabilities that can anticipate pedestrian intentions and behavioral responses in advance, thereby providing essential inputs for motion planning, risk assessment, and decision-making in autonomous vehicles. However, most existing approaches rely primarily on observable features and data-driven correlations, while lacking an explicit understanding of the cognitive mechanisms underlying pedestrian behavior.

Pedestrian safety behavior arises from a complex decision-making process rather than a single factor [2–4]. Analyses focusing on individual factors are insufficient to explain behavioral mechanisms or to support reliable prediction [5]. To address this limitation, some studies have adopted theories from human behavioral science to model the joint effects of multiple factors on pedestrian behavior. However, their reliance on surveys and questionnaires constrains the ability to capture underlying cognitive processes. Cognition is a key determinant of human behavior, and many psychological models describe decision-making as a sequence of cognitive stages [5]. Questionnaire-based methods provide only retrospective and coarse representations of these stages, limiting the interpretation of complex decision-making mechanisms. Recent advances in physiological sensing technologies offer alternative approaches by operationalizing cognitive states, such as emotion [6], attention [7], and risk perception [8], through electroencephalography and eye-tracking. These methods enable objective and time-resolved assessment of cognitive dynamics beyond self-reported measures.

In this study, multimodal data, including behavioral records, questionnaires, eye-tracking metrics, and electroencephalography signals, are integrated to construct indicators corresponding to different stages of the decision-making process. A structural equation model is then employed to examine the relationships among multiple factors and to elucidate the cognitive mechanisms underlying pedestrian safety behavior.

2 Related work

In research on pedestrian safety, a common approach involves analyzing historical accident reports and video data to examine key factors related to traffic, roadway conditions, socio-demographic characteristics and individual factors. The findings from such analyses often inform recommendations in areas such as safety education, traffic law enforcement, and road design. However, when exploring the relationship between factors and final behavior, these studies often only concentrate on single or a few variables. In reality, pedestrian's safety decision making issues are complex system problems, with various factors interrelated and interacting. In this regard, some researchers have integrated certain mental models from psychology and behavioral science, striving to generalize their application scenarios to the field of transportation and deduce theoretical models of pedestrian behavior [9]. Yet, key variables associated with interaction in complex environments and risk-related decision-making, such as attention [10] and vigilance [11], remain largely underexplored remain largely underexplored because they are inherently difficult to measure. In recent years, the emergence of physiological sensing technologies, such as eye-tracking and electroencephalography (EEG), has provided new opportunities to capture the dynamic nature of human cognition based on neuroscientific evidence. The electrical signals recorded by these devices possess high resolution and contain rich mental features, offering valuable insights into the neural and cognitive mechanisms underlying human decision-making [12,13]. When integrated with virtual simulation environments or high-fidelity experimental platforms, these technologies enable real-time monitoring of cognitive indicators such as attention allocation, risk perception, and decision response within realistic traffic contexts. This advancement not only enhances our understanding of the psychological and neural mechanisms underlying pedestrian decision-making but also contributes to the development of more precise predictive models for traffic safety.

While multiple cognitive dimensions may theoretically contribute to pedestrian decision-making, including perceived severity, vulnerability, emotional arousal, and decision confidence, not all of these constructs are equally relevant or measurable within dynamic traffic contexts. This study focuses on cognitive components that are both theoretically central to decision-making and operationalizable in time-sensitive environments. In particular, based on the Situation Awareness (SA) framework, attention-related processes (captured as visual engagement) and higher-level cognitive evaluation processes (captured as risk awareness) were selected as the primary constructs. These components directly influence how pedestrians perceive, interpret, and respond to traffic risks in real time. The selection of these constructs was therefore theory-driven rather than measurement-driven. Eye-tracking and EEG techniques were adopted not to constrain variable selection, but to provide objective and time-resolved measurements of these theoretically grounded cognitive processes.

3 Methodology

To ensure the methodological rigor and theoretical consistency of this study, the research design was developed based on the nature of pedestrian decision-making as a complex cognitive process occurring in dynamic traffic environments. Pedestrian safety behavior is not directly observable as a simple reaction to external stimuli, but rather emerges from the interaction between environmental conditions, individual characteristics, and internal cognitive mechanisms. Therefore, a multimodal and theory-driven approach was adopted to capture these heterogeneous factors and their interrelationships.

Specifically, the Situation Awareness (SA) framework [5] was selected as the theoretical foundation of the cognitive process, enabling the systematic operationalization of latent constructs. To quantitatively model the interactions among these latent variables and their observed indicators, Structural Equation Modeling (SEM) was employed, as it allows for the simultaneous estimation of measurement and structural relationships within a unified framework. In addition, multimodal data sources, including behavioral observations, questionnaires, eye-tracking, and electroencephalography (EEG), were integrated to provide complementary measurements of cognitive and behavioral processes, thereby enhancing construct validity and reducing reliance on single-source data. This methodological design ensures that both the cognitive mechanisms and observable behaviors of pedestrians can be jointly examined, providing a robust basis for uncovering the underlying pathways influencing safety performance. The overall research framework is illustrated in Figure.1. First, the latent constructs and their corresponding measurement variables were designed by conceptually adapting the abstract components of the classical Situation Awareness (SA) framework to the target scenario of this study. Second, an interactive simulation experiment focuses on the pedestrians' street-crossing scenario was conducted to collect multimodal data for metric computation. The measurement variables were obtained through an integrated approach combining behavioral observation, questionnaire surveys, and physiological sensing. The key physiological indicators were derived through signal processing procedures, including EEG feature extraction based on wavelet packet transform (WPT) and eye-tracking feature extraction based on areas of interest (AOI). Finally, the SEM was evaluated and optimized to construct the final model, and the internal mechanisms of pedestrian decision-making were analyzed.

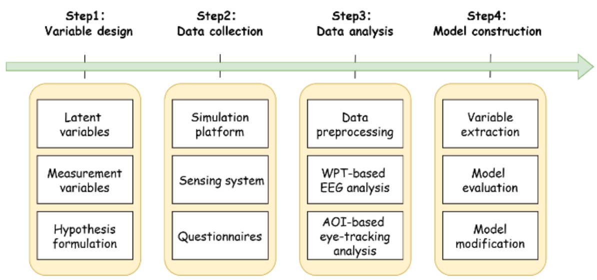

Figure.1 The overall research framework

3.1 Variable design

Pedestrian decision making in traffic settings is a typical dynamic decision task within a complex system, where perception, comprehension, and projection must be continuously updated under time pressure and uncertainty. Based on the SA framework, the key components were conceptually derived and operationalized, and the resulting latent constructs and their corresponding measurement variables are summarized in Table.1

Table.1 Variable design of the SEM

Latent Variables	Measured Variables
Hazardous Atmosphere	Complexity of vehicles (CV) Complexity of vulnerable road users (CVRU)
Safety Propensity	Risk-taking preference (RTP) Traffic rule compliance (TRC) Traffic accident sensitivity (TAS)
Visual Engagement	Total fixation proportion (TFP) AOI fixation frequency (AOIFF) AOI glance count (AOIGC) Average saccade amplitude (ASA) Maximum fixation duration (MFD) Saccade-to-fixation duration ratio (SFR)
Risk Awareness	Vigilance (VIG) Mental workload (MWL) Perceived safety performance (PSP)
Safety Performance	Environmental observation (EO) Speed regulation (SR) Traffic-light compliance (TLC)

3.1.1 Hazardous atmosphere

Pedestrian safety is influenced by surrounding traffic elements, including vehicles, non-motorized vehicles, and other pedestrians. These elements' number, speed, spacing, and variability define system complexity, affecting pedestrians' risk perception and information-processing load during street crossing [14–16]. In this study, hazardous atmosphere was conceptualized based on traffic complexity, including vehicle and VRU interactions. The two observed variables, complexity of vehicles (CV) and complexity of vulnerable road users (CVRU), were used to represent the level of environmental complexity in each experimental scenario. These indicators were derived based on the dynamic traffic conditions presented during the task. Specifically, by reviewing the task replay videos, the number, speed, spatial distribution, as well as the level of compliance with traffic rules of surrounding vehicles and VRUs were evaluated on a second-by-second basis throughout each task. These factors jointly reflect the dynamic characteristics and potential risk level of the traffic environment. Based on these time-resolved evaluations, an overall complexity value was calculated for each task by aggregating the second-by-second observations. The resulting values were then averaged, standardized across all tasks, and subsequently categorized into different levels, which were used as the measurement values for CV and CVRU.

3.1.2 Safety propensity

Safety propensity refers to an individual's tendency to engage in risky traffic behaviors, influenced by personality traits and experience [5]. For the measurement variables of the latent safety propensity, first, the classical Sensation Seeking Scale developed by Zuckerman [17] was adopted. The trait of sensation seeking, defined as the seeking of varied, novel, complex and intense sensations and experiences, and the willingness to take physical, social, legal and financial risks for the sake of such experiences, has long been recognized in the personality-psychology literature as a classic indicator of risk preference [18]. In addition, to enhance the contextual relevance in pedestrian traffic scenarios, another 2 self-report subscales were developed to measure Traffic rule compliance (TRC) and Traffic accident sensitivity (TAS), respectively. TRC reflects individuals' habitual compliance with traffic rules, while TAS assesses the influence of safety education and accident warnings on their behavioral intentions. Beyond the personality-based indicator of Sensation Seeking, these two variables capture the experience-based dimensions of Safety Propensity, as pedestrian safety behavior is shaped not only by stable risk preferences but also by prior compliance habits and exposure to safety information [1].

3.1.3 Visual engagement

The perception component of the SA model in this study was interpreted as visual engagement, referring to an individual's attention to key traffic elements such as traffic lights and vehicles. Various eye-tracking metrics have been proposed to evaluate the characteristics and efficiency of such engagement [19–22]. In this paper, visual engagement, representing attention to critical traffic elements, was measured using six eye-tracking metrics: total fixation proportion (TFP), fixation frequency (AOIFF), glance count (AOIGC), saccade amplitude (ASA), maximum fixation duration (MFD), and saccade-to-fixation duration ratio (SFR). The six eye-tracking metrics were selected to capture complementary aspects of visual attention rather than redundant information. Specifically, fixation-based metrics (TFP, AOIFF, MFD) primarily reflect attentional allocation and depth of processing, while glance-based metrics (AOIGC) capture the frequency of attentional shifts across dynamic elements. Saccade-related metrics (ASA, SFDR) further characterize visual search strategies and the efficiency of information acquisition. Together, these metrics represent different dimensions of visual engagement, including attentional focus, scanning behavior, and processing dynamics, thereby providing a more comprehensive operationalization of perceptual processes in dynamic traffic environments.

3.1.4 Risk awareness

The second and third stages of SA, comprehension and projection, are difficult to distinguish explicitly; however, both are inherently related to the allocation of cognitive resources and the integration of information during the decision-making process. This paper measured risk awareness through mental workload and vigilance. From a perspective of neuroscience, these level of mental workload and vigilance can be assessed through EEG, which captures real-time fluctuations in neural activity. Additionally, a self-designed questionnaire was developed to measure Perceived safety performance (PSP) through their self-evaluated satisfaction with the safety of their own behaviors, which also reflects participants' subjective risk awareness.

3.1.5 Safety performance

Safety performance was assessed by evaluating environmental observation, speed regulation, and traffic-light compliance. Environmental observation (EO) was characterized based on participants' active visual exploration behaviors recorded in the keyboard logs. Specifically, camera control inputs were analyzed to identify directional view adjustments during the task. The frequency of these adjustments, together with the extent of viewpoint changes, was used to reflect how actively participants scanned their surroundings. More frequent and wider-ranging view adjustments indicate a higher level of environmental monitoring. Speed regulation (SR) was evaluated based on movement operation patterns captured in the keyboard logs. The temporal sequence of movement commands was examined to identify changes in movement speed and control stability. Participants who maintained smoother and more consistent movement patterns, without abrupt or frequent speed changes, were considered to exhibit better speed regulation performance. Traffic-light compliance (TLC) was determined through frame-by-frame video review of the task process. Crossing behavior was aligned with the traffic signal status to identify whether participants initiated or continued crossing during red-light phases. The proportion of time associated with rule violations was then calculated for each participant and used as an inverse indicator of compliance level. For all three dimensions, the extracted behavioral features were subsequently transformed into comparable performance scores through a ranking procedure. These scores were then integrated to form the overall safety performance measure used in the structural equation model.

3.2 Eye-tracking and EEG metrics

In this study, the eye-tracking and EEG data segments for each task were extracted to define the analysis windows. The raw gaze signals were preprocessed in Tobii Pro Lab to classify fixations and saccades based on eye movement velocity (<30°/s as fixations). Temporal gaps up to 75ms were interpolated, and high-frequency noise was attenuated using a 3-sample moving median filter. The gaze data from both eyes were combined using a mean-combined metric. Fixations with a duration <60ms were excluded. Dynamic areas of interest (AOIs), including key traffic elements such as vehicles, pedestrians, and traffic lights, were manually defined. Eye-tracking metrics including fixation, glance, and saccade data were generated and used to calculate corresponding indicators.

Vigilance (VIG) and mental workload (MWL) were measured using relative beta and theta band power in prefrontal (FP1, FP2) and frontal (F3, F4) EEG channels, with baseline data as a reference. Data preprocessing included resampling to match eye-tracking data, band-pass filtering (0.16–43 Hz), notch filtering to remove power-line interference, and applying Independent Component Analysis (ICA) to remove artifacts. A baseline segment was recorded before the task and underwent the same preprocessing. The wavelet packet transform (WPT) method was used to decompose EEG signals into frequency sub-bands, with a 5-level decomposition using the Daubechies 8 (DB8) wavelet. The WPT can be expressed as:

$$d_{j+1,2n}(k) = \sum_m g(m-2k) d_{j,n}(m)$$

$$d_{j+1,2n+1}(k) = \sum_m h(m-2k) d_{j,n}(m)$$

where $d_{j,n}(k)$ denotes the wavelet packet coefficient at decomposition level j and node n ; $g(m)$ and $h(m)$ represent the low-pass and high-pass conjugate quadrature filters, respectively.

Due to differences in sampling rates across data sources, all physiological and behavioral data were segmented into task-based time windows. Within each window, features were aggregated (e.g., mean, proportion, or frequency) to obtain a single representative value per variable per trial, ensuring compatibility for SEM analysis.

3.3 Structural equation model

Structural Equation Modeling (SEM) was used to identify the set of parameters that best fit the sample covariance matrix, uncovering the underlying causal pathways and interrelationships among variables. SEM consists of two components: the measurement model and the structural model. The measurement model defines the relationships between latent variables and their observed indicators, while the structural model describes the causal relationships among the latent variables.

To address the nested data structure resulting from multiple trials per participant, Maximum Likelihood estimation with Robust standard errors (MLR) was employed, with the corresponding log-likelihood function defined as:

$$\log L(\theta) = -\frac{1}{2}\sum_{i=1}^{N}[\log|\Sigma_i| + (y_i - \mu_i)^T \Sigma_i^{-1}(y_i - \mu_i)]$$

where Σ_i represents the implied covariance matrix predicted from the parameter vector θ for the i^{th} observation, y_i is the observed data vector, μ_i is the model-implied mean vector, and N is the total sample size. The objective function of the MLR estimation is given by:

$$F_{ML} = log|\Sigma(\theta)| + tr\left(S\Sigma^{-1}(\theta)\right) - log|S| - p$$

where S is the sample covariance matrix and p is the number of observed variables. This method accounts for the non-independence of observations and provides more accurate estimates in the presence of clustering. The model parameters were iteratively estimated using the Newton-Raphson algorithm, and clustered standard errors were adjusted using the Huber-White Sandwich Estimator to account for the hierarchical nature of the data.

4 Experiment

In this paper, a street-crossing scenario within an urban environment with traffic lights was developed using the Python API from the CARLA simulator and visual rendering through Unreal Engine 4. Participants explored the environment by adjusting their field of view and moving using keyboard input. The task required participants to comply with traffic lights, observe their surroundings, and cross the road along the pedestrian lane. To simulate real-world conditions, the scene included urban infrastructure, moving vehicles, and vulnerable road users (VRUs), including pedestrians and cyclists, as shown in Figure 3. 30 male participants from Beijing, China, participated in the experiment, all holding valid driver's licenses and understanding traffic regulations. Before the experiment, they completed questionnaires assessing their traffic behaviors and sensation-seeking tendencies. Traffic behavior and safety awareness were also assessed through a custom questionnaire. The experimental simulation was displayed on a 27-inch monitor, with eye movements recorded by a Tobii Pro Fusion eye tracker at 125 Hz. EEG data were recorded using a 32-channel Emotiv FLEX wireless system, sampled at 128 Hz, and transmitted via Bluetooth. Both eye-tracking and EEG data were synchronized and recorded throughout the experiment. The experimental procedure included questionnaire completion, equipment fitting, environment familiarization, eye-tracking calibration, EEG baseline recording, and subsequent tasks followed by questionnaires. Data quality was monitored in real-time through Tobii Pro Lab and Emotiv Pro, ensuring consistent signal acquisition. The software and hardware configuration of the equipment is shown in Figure 2.

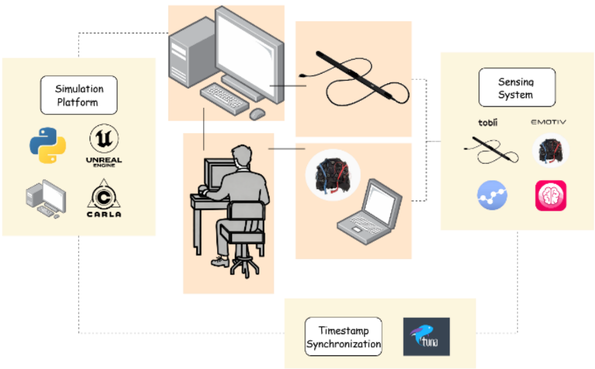

Figure 2. Layout of the experiment settings

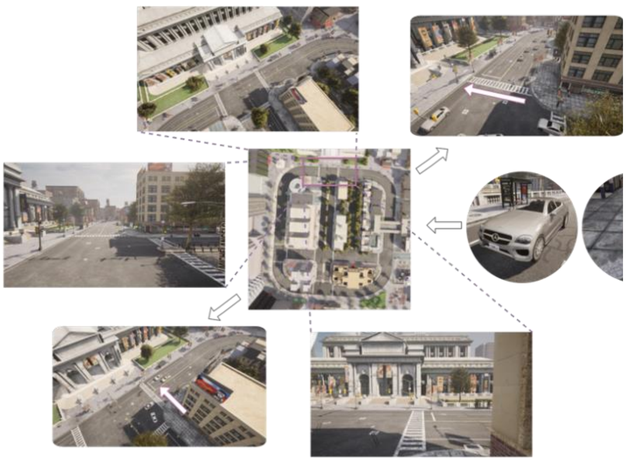

Figure 3. Environment of the experiment task

5 Result and discussion

The measurement model was evaluated through Confirmatory Factor Analysis (CFA), and the results indicated good validity and internal consistency. The model fit was assessed using several commonly adopted indices. The the chi-square to degrees of freedom ratio (χ^2/df) reflects overall model fit, with values close to 1 indicating good fit. CFI and TLI assess incremental fit compared to a null model, with values above 0.95 considered excellent. RMSEA evaluates approximation error, with values below 0.05 indicating close fit, while SRMR reflects residual differences, with values below 0.08 indicating acceptable fit. Finally, the χ^2/df was 1.32, CFI was 0.979, TLI was 0.974, RMSEA was 0.037, and SRMR was 0.066. The model demonstrates an excellent fit. The path diagram of model is shown in Figure 4. All measurement indicators showed significant loadings on their respective latent constructs, with standardized factor loadings ranging from 0.715 to 0.909 for Safety Propensity, 0.726 to 0.855 for Visual Engagement, 0.814 to 0.863 for Risk Awareness, and 0.781 to 0.84 for Safety Performance, demonstrating strong convergent validity. Hazardous Atmosphere significantly influenced both Visual Engagement ($\beta = 0.521$, $p < 0.05$) and Risk Awareness ($\beta = 0.557$, $p < 0.01$), suggesting that higher environmental complexity and reduced visibility enhance pedestrians' attention and risk perception. Additionally, Safety Propensity positively affected Risk Awareness ($\beta = 0.26$, $p < 0.05$), indicating that individuals with a higher safety orientation are more attuned to potential hazards in traffic. Furthermore, Risk Awareness had a significant and robust impact on Safety Performance ($\beta = 0.611$, $p < 0.001$), supporting its role as a mediator between cognitive perception and behavioral outcomes.

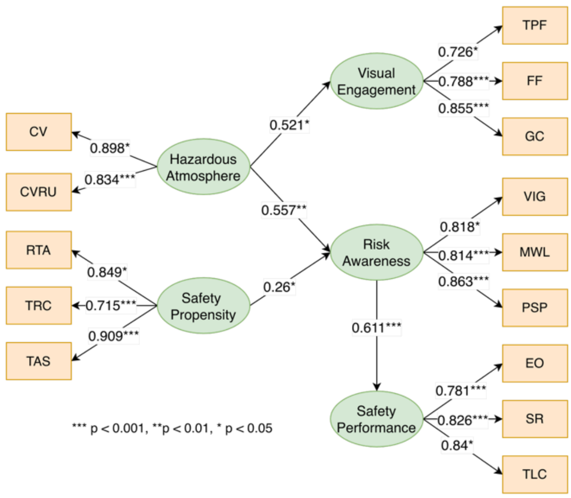

Figure 4. SEM path diagram

The findings underscore the central importance of cognitive mechanisms in shaping pedestrian safety behavior. The effects of Hazardous Atmosphere and Safety Propensity on Safety Performance were fully mediated by cognitive constructs, particularly Risk Awareness. This suggests that pedestrians do not directly translate environmental hazards or personal safety tendencies into behavioral responses. Instead, these influences must first be cognitively appraised and integrated before manifesting as observable safety behaviors. From a theoretical standpoint, this aligns with cognitive human factors research, which emphasizes that risk-related decisions in dynamic environments are critically dependent on internal cognitive evaluations rather than external stimuli alone.

A key observation is that Visual Engagement did not have a statistically significant direct effect on either risk awareness or safety performance. This suggests that the influence of visual attention on behavior is not straightforward. Visual perception likely plays an indirect role, shaping cognitive constructs like risk awareness, rather than directly influencing behavior. This interpretation is supported by theories in visual cognition, which distinguish between low-level perceptual attention (where individuals look) and high-level cognitive attention (how visual information is interpreted). Therefore, perceptual attention alone is insufficient without subsequent cognitive processing, which highlights the layered nature of visual cognition.

The results also indicate that visual engagement, when concentrated on specific elements such as traffic signals or vehicles, does not necessarily enhance risk awareness. In contrast, broader visual scanning patterns seem to reinforce risk perception, suggesting that excessive focus on narrow visual cues could limit awareness of broader environmental threats. A more holistic visual strategy helps pedestrians form a more comprehensive mental model of environmental risks. These findings have important implications for traffic safety interventions: designing road environments to facilitate broader and more efficient visual scanning, reducing visual clutter, and structuring cues to encourage wider scanning ranges could improve both risk awareness and safety performance.Furthermore, the non-significant direct effect of Visual Engagement suggests the importance of selecting appropriate measurement indicators. The relationship between the breadth and depth of visual perception is complex, with breadth-oriented indicators capturing general scanning behaviors and depth-oriented indicators reflecting the processing efficiency of risk-relevant cues. Pedestrian decision-making depends more on depth-oriented measures, making it crucial to choose eye-tracking indicators that align with the perceptual demands specific to traffic safety contexts.

The comparison between Safety Propensity and Hazardous Atmosphere also reveals that visual engagement is more strongly influenced by situational hazards than by personality traits. This suggests that visual attention in traffic settings is primarily stimulus-driven rather than trait-driven. Pedestrians often lack standardized visual strategies, unlike professional drivers who are trained in visual scanning and hazard detection. This highlights a practical implication: pedestrian safety interventions should extend beyond awareness campaigns to include targeted training aimed at enhancing perceptual skills and hazard recognition abilities.

Finally, Risk Awareness was identified as a dominant predictor of Safety Performance, with pedestrians in more hazardous environments demonstrating stronger feelings of insecurity, which in turn promote safer behavioral choices. However, the stronger influence of Hazardous Atmosphere on Risk Awareness compared to Safety Propensity emphasizes the dominant role of situational factors. Environmental hazards elicit immediate cognitive and emotional responses, whereas personality traits exert slower, more diffuse influences. This suggests that improving pedestrian safety requires not only behavioral interventions but also environmental design strategies that reduce risk cues or communicate hazards more effectively. Enhancing lighting, improving sightlines, and optimizing pedestrian signage can help reduce perceived insecurity and promote safer decisions.

6 Conclusion

In conclusion, this study presents an integrated framework for investigating the mechanisms of pedestrians' street-crossing behavior using structural equation modeling and multimodal data. By embedding these indicators into a Situation Awareness-based model, the study provides a comprehensive understanding of how environmental, individual, perceptual, and cognitive factors shape pedestrian safety performance. The findings offer both theoretical insights and practical implications for traffic safety management.

Several limitations should be noted in this study. To ensure participant safety, ethical compliance, and sufficient experimental control—especially when investigating risky behaviors such as unsafe street crossing—all experiments were conducted in a virtual traffic environment rather than in real-world traffic scenarios. Although the CARLA platform provides realistic urban layouts, dynamic traffic interactions, and high ecological validity, and although existing literature supports that physiological responses in high-fidelity virtual environments can closely reflect cognitive processes comparable to those in real conditions, virtual settings still cannot fully replicate the realistic risk perception, emotional arousal, and decision pressure that individuals experience in actual traffic environments.

To address these constraints and further advance the research, several promising directions can be pursued in future work. First, researchers may conduct hybrid validation studies that combine virtual experiments with small-scale, safely managed real-world tests to improve the external validity of findings. Second, future studies could incorporate more immersive sensing or feedback mechanisms to enhance the realism of risk perception and emotional engagement in virtual environments. Third, the models and findings derived from simulation could be further deployed and tested in real-world edge computing or intelligent infrastructure systems to support practical applications such as risky behavior warning, pedestrian protection, and smart crossing assistance.

7 Acknowledgment

This work was financially supported by Mercedes-Benz Group China Ltd. -Tsinghua Joint Research Institute (#04191400124/20223910001). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of Mercedes-Benz Group China Ltd.

References

[1] An extension of the theory of planned behavior to predict pedestrians' violating crossing behavior using structural equation modeling, Accident Analysis & Prevention 95 (2016) 417–424. https://doi.org/10.1016/j.aap.2015.09.009.

[2] A. Tom, M.-A. Granie, Gender differences in pedestrian rule compliance and visual search at signalized and unsignalized crossroads, Accid. Anal. Prev. 43 (2011) 1794–1801. https://doi.org/10.1016/j.aap.2011.04.012.

[3] K. Aghabayk, S. Rejali, N. Shiwakoti, The Role of Big Five Personality Traits in Explaining Pedestrian Anger Expression, Sustainability 14 (2022) 12099. https://doi.org/10.3390/su141912099.

[4] R. Lobjois, V. Cavallo, Age-related differences in street-crossing decisions: The effects of vehicle speed and time constraints on gap selection in an estimation task, Accid. Anal. Prev. 39 (2007) 934–943. https://doi.org/10.1016/j.aap.2006.12.013.

[5] M. Endsley, Toward a Theory of Situation Awareness in Dynamic-Systems, Hum. Factors 37 (1995) 32–64. https://doi.org/10.1518/001872095779049543.

[6] P. Mavros, M.Z. Austwick, A.H. Smith, Geo-EEG: Towards the Use of EEG in the Study of Urban Behaviour, Appl. Spat. Anal. Policy 9 (2016) 191–212. https://doi.org/10.1007/s12061-015-9181-z.

[7] S. Tang, J. Wang, W. Liu, Y. Tian, Z. Ma, G. He, H. Yang, A study of the cognitive process of pedestrian avoidance behavior based on synchronous EEG and eye movement detection, Heliyon 9 (2023) e13788. https://doi.org/10.1016/j.heliyon.2023.e13788.

[8] H. Lee, O. Lim, A. Singh, S. Samuel, Collision Risk Perception Models Using Physiological and Eye-Tracking Signals, IEEE Access 13 (2025) 136745–136767. https://doi.org/10.1109/ACCESS.2025.3594806.

[9] Pedestrian crossing decision-making: A situational and behavioral approach, Safety Science 47 (2009) 1248–1253. https://doi.org/10.1016/j.ssci.2009.03.016.

[10] N. Valos, J.M. Bennett, The relationship between cognitive functioning and street-crossing behaviours in adults: A systematic review and meta-analysis, Transp. Res. Pt. F-Traffic Psychol. Behav. 99 (2023) 356–373. https://doi.org/10.1016/j.trf.2023.10.018.

[11] I. Pless, H. Taylor, L. Arsenault, The Relationship Between Vigilance Deficits and Traffic Injuries Involving Children, Pediatrics 95 (1995) 219–224.

[12] W. Klimesch, EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis, Brain Res. Rev. 29 (1999) 169–195. https://doi.org/10.1016/S0165-0173(98)00056-3.

[13] K. Rayner, Eye movements and attention in reading, scene perception, and visual search, Q. J. Exp. Psychol. 62 (2009) 1457–1506. https://doi.org/10.1080/17470210902816461.

[14] A. Rasouli, I. Kotseruba, J.K. Tsotsos, Understanding Pedestrian Behavior in Complex Traffic Scenes, IEEE T. Intell. Veh. 3 (2018) 61–70. https://doi.org/10.1109/TIV.2017.2788193.

[15] B.R. Kadali, P. Vedagiri, N. Rathi, Models for pedestrian gap acceptance behaviour analysis at unprotected mid-block crosswalks under mixed traffic conditions, Transp. Res. Pt. F-Traffic Psychol. Behav. 32 (2015) 114–126. https://doi.org/10.1016/j.trf.2015.05.006.

[16] J.J. Faria, S. Krause, J. Krause, Collective behavior in road crossing pedestrians: the role of social information, Behav. Ecol. 21 (2010) 1236–1242. https://doi.org/10.1093/beheco/arq141.

[17] M. Zuckerman, D. Mangelsdorff, R. Neary, B. Brustman, R. Bone, What Is Sensation Seeker - Personality Trait and Experience Correlates of Sensation-Seeking Scales, J. Consult. Clin. Psychol. 39 (1972) 308-+. https://doi.org/10.1037/h0033398.

[18] T. Zheng, W. Qu, Y. Ge, X. Sun, K. Zhang, The joint effect of personality traits and perceived stress on pedestrian behavior in a Chinese sample, PLoS One 12 (2017) e0188153. https://doi.org/10.1371/journal.pone.0188153.

[19] Q. Zhao, X. Zhuang, T. Zhang, Y. He, G. Ma, Pedestrian gaze pattern before crossing road in a naturalistic traffic setting, Eur. Transp. Res. Rev. 15 (2023) 31. https://doi.org/10.1186/s12544-023-00605-1.

[20] F. Biassoni, M. Bina, F. Confalonieri, R. Ciceri, Visual exploration of pedestrian crossings by adults and children: Comparison of strategies, Transp. Res. Pt. F-Traffic Psychol. Behav. 56 (2018) 227–235. https://doi.org/10.1016/j.trf.2018.04.009.

[21] J. de Winter, P. Bazilinskyy, D. Wesdorp, V. de Vlam, B. Hopmans, J. Visscher, D. Dodou, How do pedestrians distribute their visual attention when walking through a parking garage? An eye-tracking study, Ergonomics 64 (2021) 793–805. https://doi.org/10.1080/00140139.2020.1862310.

[22] H.E. Clark, J.A. Perrone, R.B. Isler, S.G. Charlton, The role of eye movements in the size-speed illusion of approaching trains, Accid. Anal. Prev. 86 (2016) 146–154. https://doi.org/10.1016/j.aap.2015.10.028.