



“G. D'Annunzio” SCHOOL OF ADVANCED STUDIES

Ph.D. in BUSINESS & BEHAVIORAL SCIENCES

XXXVIII CYCLE

**Human Performance in Safety-Critical Scenarios:
A Conceptual and Empirical Framework for Assessing
Perception, Confidence Calibration, and Risk Estimation**

Department of Neuroscience, Imaging and Clinical Sciences

Scientific-Disciplinary Sector: M-PSI/01, SECS-P/07

PhD Student
Dott.ssa Elena Lupia

Coordinator & Tutor
Prof. Riccardo Palumbo

Academic Years 2022–2025

Abstract

Understanding human performance in safety-critical scenarios needs a multi-component perspective that accounts for the interaction between cognitive processes, metacognitive monitoring, task demands and environmental factors. Methodologically, this thesis combined a systematic review and two empirical studies to investigate how visual information is acquired, how confidence and risk are evaluated, and how these evaluations manifest across perception and decision-making tasks in dynamic scenarios. The systematic review consolidated evidence on differences in gaze behavior between experts and novices during perceptual processing and on the potential for gaze-based skill transfer. It showed that expert gaze tends to be more selective, stable, and anticipatory, and that the effectiveness of gaze-based training depends on task complexity and learners' prior knowledge. The second study introduced a novel adaptation of the Hazard Perception Task, which integrated trial-by-trial confidence judgments and risk estimation as measures of metacognitive monitoring. Empirically it showed that overconfidence impaired (driving) hazard detection accuracy without affecting reaction times. Risk estimation was shaped by environmental factors and was positively correlated with environmental sensitivity. The third study examined confidence and risk estimation across the (driving) Hazard Perception Task and the (cycling) Decision-Making Task, to assess metacognitive consistency. Confidence and risk judgments showed strong cross-task correlations, while task performance was more sensitive to situational factors. Consistency across different task demands and stimulus types suggests that this reflects a generalizable metacognitive tendency rather than stimulus-specific effects. Across the empirical studies, the traffic and mobility domain is adopted as an applied paradigmatic case of complexity. However, the constructs explored are broadly applicable to other safety-critical contexts, including aviation, healthcare, and maritime operations.

Contents

Abstract	1
Contents	2
List of Tables	5
List of Figures	6
Acronyms	8
1. Introduction	9
1.1 Framework Overview	9
1.2 Thesis Contributions	12
1.3 Outline of the Thesis	14
1.4 Publications	15
References	16
2. Skill Transfer and Gaze Strategies Studied by Eye-Tracking:	
A Systematic Review	22
Abstract	23
2.1 Introduction	24
2.1.1 Objectives	25
2.2 Conceptual Background	26
2.2.1 Skill Transfer Approaches	26
2.2.2 Gaze-based Training Interventions: Application Fields	28
2.2.3 Eye-Tracking Metrics	30
2.3 Methodology	33
2.3.1 Literature Search and Selection Strategy	33
2.3.2 Descriptive Analysis	35
2.4 Key findings	36
2.4.1 Expertise-Related Differences in Perceptual Processing	37
2.4.2 Effects of Gaze-based Interventions on Task Performance	38

2.5 Discussion	40
2.5.1 Limitations and Implications	43
2.6 Conclusion	44
References	45
3. The Impact of Overconfidence and Environmental Conditions on Hazard Perception and Risk Assessment: An Experimental Study Using Video-Based Traffic Scenarios	61
Abstract	62
3.1 Introduction	63
3.1.1 The role of Confidence in Hazard Perception	64
3.1.2 Individual differences in Environmental Sensitivity	66
3.1.3 Overconfidence and Driving Experience	67
3.1.4 The current study	67
3.2 Method	68
3.2.1 Participants	68
3.2.2 Materials	68
3.2.3 Experimental Design	71
3.2.4 Procedure	72
3.2.5 Data Analysis	72
3.3 Results	74
3.3.1 Descriptive statistics	74
3.3.2 Behavioral data analysis	75
3.4 Discussion	79
3.4.1 Limitations and Implications	81
3.5 Conclusion	81
References	83
4. Confidence and Risk in Hazard Perception and Dynamic Decision Making in Traffic	96
Abstract	97

4.1 Introduction	98
4.1.1 Perception and Decision Making in Traffic	99
4.1.2 The Role of Metacognition in Traffic	100
4.1.3 Study Aims and Hypotheses	100
4.2 Method	101
4.2.1 Participants	101
4.2.2 Materials	102
4.2.3 Design	105
4.2.4 Procedure	107
4.2.5 (Statistical) Analysis	107
4.3 Results	109
4.3.1 Hazard Perception Task	109
4.3.2 Decision-Making Task	112
4.3.3 Correlation Analyses between Hazard Perception and Decision-Making	113
4.4 Discussion	117
4.4.1 Limitations and Implications	118
4.5 Conclusion	119
References	120
5. General Discussion	129
5.1 Summary of the Main Findings	129
5.2 Theoretical Implications	131
5.3 Practical Implications	131
5.4 Limitations and Future Directions	132
References	134
Supplementary Materials Section A - Experiments Materials	136
Supplementary Materials Section B - Full Model Estimates	158

List of Tables

3.1 Comparison of descriptive statistics between groups. (p. 75)

3.2 Model based effects relative to the safe-daytime baseline (except for confidence bias, which reflects the main effect of the continuous predictor). Findings are reported as percentage point (pp) difference for accuracy, milliseconds for RTs, and (Likert) point differences for subjective risk estimation. Significance: *: $p < .05$; **: $p < .01$; ***: $p < .001$. Dashes (—) indicate effects not estimated. Downward arrow (↓) indicates a negative effect. (p. 78)

4.1 Predictors of hazard detection accuracy, response times, risk estimates, and confidence judgments, pp = percentage points; s = seconds; ns = non-significant. Significance: *: $p < .05$; **: $p < .01$; ***: $p < .001$. (p. 111)

4.2 Predictors of braking, response times, risk estimates, and confidence. Notes. pp = percentage point; points = raw scale points; s = seconds. ns = non-significant. Significance: *: $p < .05$; **: $p < .01$; ***: $p < .001$. (p.113)

List of Figures

- 1.1.** Model of situation awareness proposed by Endsley (1995). (p. 11).
- 2.1** Literature selection flow diagram based on the PRISMA approach. (p. 34)
- 2.2** VOSviewer visualization of a term co-occurrence network. (p. 36)
- 3.1** Example of a single trial of the task. The timeline shows: the initial mask slide (duration 350 ms), followed by a road traffic video clip (approximately 8 seconds) during which participants could respond or not; after each video, they provided trial-by-trial rating of confidence judgment and subjective risk estimation on a scale from 1 to 7. (p. 71)
- 3.2.** Model-based estimates from mixed effects model as a function of hazard presence and lighting conditions (95% CIs). Panels show: (A) predicted accuracy from the generalized linear mixed-effects model (GLMM), and (B) estimated marginal means of risk ratings from the linear mixed-effects model (LMM). (p. 77)
- 4.1** Sequence of events in a trial of the Hazard Perception Task. (p. 104)
- 4.2** Sequence of events in a trial of the Decision-Making Task. (p. 104)
- 4.3** Example frames from the Hazard Perception Task, showing clips containing hazards (A, C) and without hazardous events (B, D), presented during daytime (A, B) and nighttime (C, D) conditions. (p. 105)
- 4.4** Example screenshots from scenarios of the Decision-Making Task in areas with high (A, C) and low density (B, D) and representing different environments (A, B at the Janskerkhof, and C, D at the Stadhuisplein). (p. 106)

4.5. Marginal density plots display distributions between (Driving) Hazard Perception Task and (Cycling) Decision-Making Task. Panels show A) performance: accuracy vs. braking, B) mean risk estimation, and C) mean confidence judgments. (p.115)

4.6. Marginal density plots display distributions between (Driving) Hazard Perception Task and (Cycling) Decision-Making Task. Panels show A) mean risk estimation in hazard trials, B) mean risk estimation in no hazard trials, C) mean confidence judgments in hazard trials, and D) mean confidence judgments in no hazard trials. (p.116)

Acronyms

HF Human Factors

SA Situational Awareness

EMME Eye Movement Modeling Examples

QE Quiet Eye

QET Quiet Eye Training

AOIs Areas of Interest

MWL Mental Workload

CBCVIs Catheter-based Cardiovascular Interventions

HPT Hazard Perception Task

HPQ Hazard Perception Questionnaire

HSP Highly Sensitive Person scale

HPS Hazard Perception Skill

BART Balloon Analogue Risk Task

RTs Reaction Times

GLMMs Generalized linear mixed-effects models

LMMs Linear Mixed Models

JOP Judgment of Performance

CRs Confidence Ratings

1

Introduction

“I looked, but I didn’t see it” captures a recurring failure in safety-critical performance, where relevant information is visually available but not effectively perceived or acted upon (Castro, 2008). This thesis builds on this insight to frame human performance as the outcome of interacting perceptual, metacognitive monitoring, and decision-making processes operating under dynamic task demands and environmental factors.

Chapter 1 briefly introduces this framework by outlining how visual information is perceived, how the accuracy of one's perceptual performance is evaluated and monitored, and how these processes jointly guide decision-making in dynamic, complex scenarios.

1.1 Framework Overview

Human performance can be understood as the observable outcome of a dynamic human-system interaction (Proctor & Van Zandt, 2018). The field of Human Factors (HF) provides a framework for understanding the interactions among humans and other elements of a system (International Ergonomics Association, 2000), with the aim to increase safety and system performance (Salvendy, 2012; Wickens et al., 2021). Within this broader field, Engineering Psychology is grounded in principles and empirical findings from cognitive psychology, and focuses on how people perceive, allocate attention, encode information and make decisions (Wickens et al., 2021).

In high-risk work environments, such as road transportation (Castro, 2008; Huang et al., 2025), aviation (Wise et al., 2016; Behrend & Dehais, 2020), maritime (Grech et al., 2019), and construction (Zhang et al., 2023), individuals operate in complex dynamic systems in which decisions require continuous adaptation to change in task demands and environmental factors. As such, relevant information is transient and embedded in constantly evolving visual scenes, constantly challenging perceptual and attentional processes (Castro, 2008). Such industries are further characterized, for instance, by information overload, time pressure, unpredictable events (e.g. sudden hazards), and uncertainty (Reale et al., 2023; Arnold et al., 2023).

The ability to “know what is going on” in a situation, that is to maintain an accurate and timely understanding of the current state of a dynamic scenario, is critical to performance in safety-critical environments (Endsley, 1995, 2017). A key concept in HF for characterizing this ability is Situational Awareness (SA), which refers to the “perception of environmental elements within a volume of time and spatial, understanding their meaning and projecting their status in the near future” (Endsley, 1995). Endsley (1995) articulated SA as a three-level model in which Level 1 involves the perception of relevant environmental cues, Level 2 involves comprehending the current situation by integrating perceived cues with pre-existing knowledge and goals, and Level 3 involves projecting how the situation will evolve in the future given current dynamics (see Figure 1.1). High levels of SA are a fundamental requirement for effective performance in these environments (Endsley, 2015). Failures at any of these levels of SA, for example not recognizing a hazard or misjudging its implications, can compromise hazard perception and decision-making, and thus may increase the likelihood of negative outcomes (Stanton et al., 2001; Grech et al., 2002; Salmon et al., 2016). Recent systematic review links human error and impaired situational awareness to accidents spanning eleven major domains, including aviation, driving, power systems, healthcare and medicine, and emergency management (see Fang et al., 2025).

Developing and maintaining situation awareness in a dynamic environment hinges also on the individual’s difference in expertise (Endsley, 2018; Cak et al., 2020). With experience, individuals develop internal models of the systems in which they operate that help guide attention (e.g., extracting meaningful cues from complex visual

scenes, and recognizing subtle patterns), reduce cognitive load, and enable anticipation of unfolding events (Endsley & Garland, 2000). This performance gap stems in part from differences in visual attention strategies, what experts choose to look at, when, and for how long. Eye tracking research has provided clear evidence that expert operators exhibit more selective and efficient task-relevant gaze patterns relevant to the task in all fields, so capturing expert gaze and use it as a basis for training novices has attracted increasing attention in applied cognitive research (for reviews see Ashraf et al., 2018; Martinez-Marquez et al., 2021).

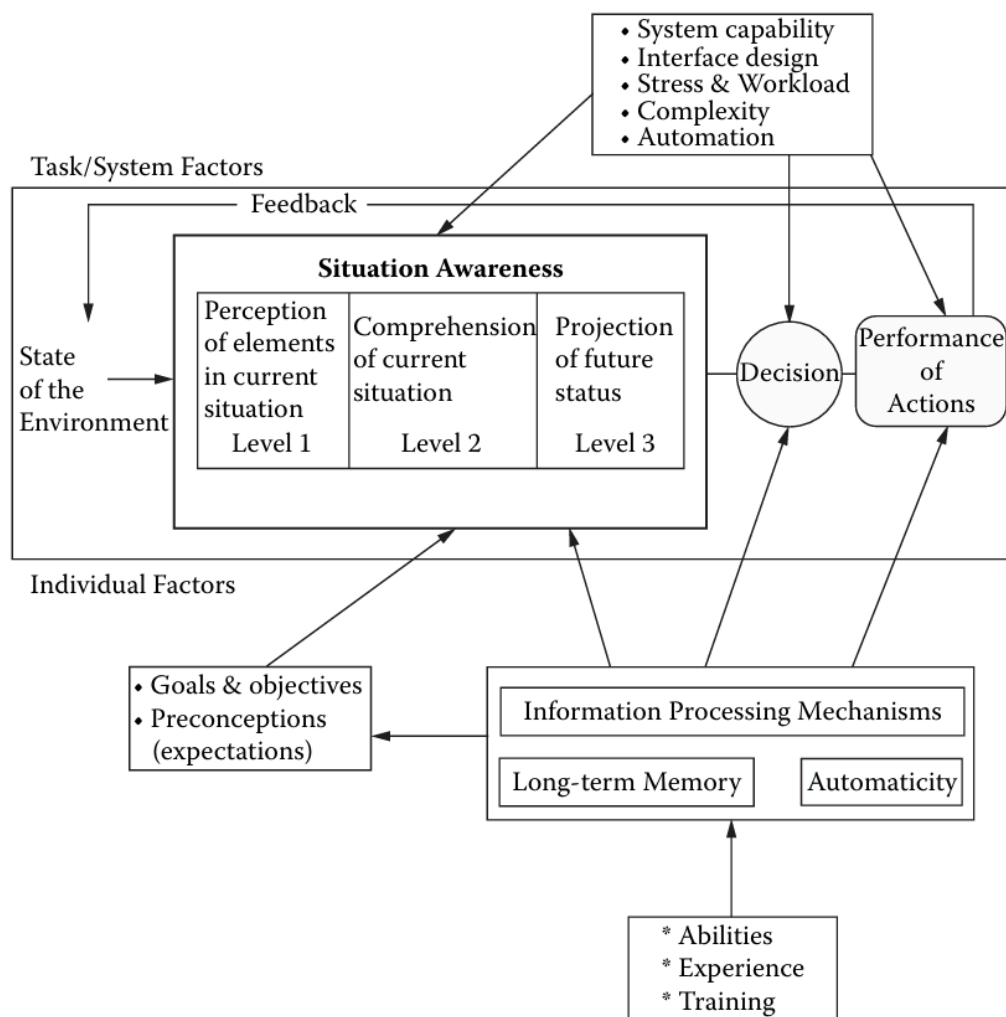


Figure 1.1. Model of situation awareness proposed by Endsley (1995).

Beyond first-order cognitive processes and experience supporting human performance and situation awareness, metacognitive higher-order mechanisms play a

critical role for self-regulation and monitoring the reliability of one's own behavior and actions (Veenman et al., 2006; Koriat, 2007; McAnally et al., 2017). An important manifestation of metacognition is the confidence judgment, defined as a subjective estimate of the likelihood of a chosen decision's success (Yeung & Summerfield, 2012; Kepecs & Mainen, 2012; Fleming & Lau, 2014; Fetsch et al., 2014). According to Koriat's (1997) cue-utilization approach, such judgments are not based on an internal "truth meter", but are inferential assessments constructed from multiple cues available to the individual at the time of judgment, such as perceived task difficulty, and familiarity. Because these cues vary in their diagnostic validity, confidence may diverge from actual performance. As a consequence, a phenomenon known as miscalibration may occur, leading to discrepancies between subjective belief and objective accuracy performance (Koriat, 1997, 2007). When confidence exceeds actual accuracy, overconfidence arises (Moore & Healy, 2008), a bias that has been shown to distort judgments and choices in a wide range of decision-making contexts (Gilbey et al., 2021), such as road safety (Wohleber & Matthews, 2016). Horrey and colleagues (2015) proposed a conceptual framework linking skill and judgment calibration to adaptive driving performance, demonstrating that drivers often misjudge both their own abilities and the demands of the situation.

In the context presented above, there is a clear need to integrate cognitive processes, metacognitive monitoring and environmental factors, within a unified framework to better understand their combined relevance on human performance.

1.2 Thesis Contributions

This thesis contributes to the behavioral sciences and applied human factors literature by advancing a multi-component understanding of how expertise, metacognitive monitoring (confidence judgments, risk estimation), and environmental factors are relevant to human performance in safety-critical scenarios. By drawing from cognitive psychology perspectives, this thesis provides different contributions linking a systematic review (Chapter II) with two empirical studies (Chapters III and IV).

First, through a systematic literature review, it consolidates existing knowledge regarding gaze behavior strategies that differentiate experts from novices and evaluates

the potential for gaze-based skill transfer across different fields. By means of this, it anchors the concept of perceptual performance as a measurable, transferable construct linked to visual attention strategies.

Second, it introduces a novel performance-based approach for measuring overconfidence in hazard perception task, tested across manipulated environmental factors (i.e., daytime vs. nighttime). It provides an integrated framework for examining perceptual performance alongside confidence judgments and risk estimation, extending traditional hazard perception approaches. By doing so, it contributes to a more comprehensive understanding of the mechanisms underlying perceptual performance in dynamic and potentially hazardous environments.

Third, it examines how confidence judgments and risk estimation correlate across perceptual and decision-making tasks within different mobility modes and task demands. By adopting a cross-task perspective, this work contributes to a deeper understanding of metacognitive consistency in traffic contexts, clarifying whether these judgments are trait-like or context-specific, and informs the development of assessment approaches that jointly capture perceptual performance and metacognitive calibration in traffic safety contexts.

Across the empirical studies of this thesis, the traffic and mobility domain is adopted as an applied paradigmatic case of complexity. It offers an example of safety-critical scenarios, and shifting environmental cues where perceptual, metacognitive, and behavioral processes need to operate in tight coordination. Although the studies are mobility-based, the constructs explored, such as experience, confidence, and risk estimation, are generalizable to other domains involving time pressure, dynamic information, and safety constraints. Thus, the traffic setting is employed as a controlled, ecologically valid framework for investigating core cognitive and metacognitive mechanisms of human performance.

Taken together, these contributions define a unified research aim grounded in the idea that effective human performance depends not only on the processing of visual information, but also on how individuals monitor and evaluate the adequacy of their own perception. In line with the notion captured by “I looked, but I didn’t see it”,

people form beliefs about whether they have seen enough, how confident they are in what they have perceived, and how risky a situation appears in light of the available information. These evaluative processes are tightly intertwined with perceptual processing and play a central role in regulating the selection of task-relevant information, deciding when to act, and how cautiously they act accordingly.

1.3 Outline of the Thesis

The thesis is organised in five chapters, plus introduction.

Chapter 2 synthesizes cross-domain eye-tracking evidence comparing expert and novice gaze strategies and evaluating gaze-based training approaches. It investigates how domain-specific expertise shapes visual attention across different application fields, it proposes a taxonomy of diagnostic gaze features, and practical design principles for gaze training in supporting task performance and skill acquisition.

Chapter 3 introduces the concept of calibration of confidence judgments and reports the experimental study investigating the impact of overconfidence and environmental factors on hazard perception and risk estimation of first-perspective driving scenarios.

Chapter 4, through an international empirical study, extends the analysis by exploring the generalizability of confidence judgments and risk estimation across perceptual and decision-making tasks. It highlights the concept that metacognitive processes are not task-specific, but function as transversal constructs that affect behavior across different mobility contexts (i.e., driving and cycling).

Chapter 5 provides a general discussion that integrates a summary of the main findings across the three studies, examines theoretical and practical implications, and outlines the overarching limitations of the thesis along with directions for future research.

Finally, the Supplementary Materials contains experimental materials (Section A), and additional full models estimates from the statistical analysis (Section B).

1.4 Publications

Part of the material presented in this thesis has appeared in some publications or has been submitted for publication:

- Lupia, E., Bortolotti, A., Palumbo, R., (2024) Skill Transfer and Gaze Strategies Studied by Eye-Tracking: A Systematic Review. *[Under Review to Trends in Psychology]*.
- Lupia, E., Bortolotti, A., & Palumbo, R. (2025). The Impact of Overconfidence and Environmental Conditions on Hazard Perception and Risk Assessment: An Experimental Study Using Video-Based Traffic Scenarios. *[Pre-print Submitted for Review to Cognitive Processing]*.
- Lupia, E., Bontje, F., Bortolotti, A., Palumbo, R., Janssen, C, P. (2025). Confidence and Risk in Hazard Perception and Dynamic Decision Making in Traffic. *[Under Review to Human Factors]*.

References

- Arnold, M., Goldschmitt, M., & Rigotti, T. (2023). Dealing with information overload: A comprehensive review. *Frontiers in Psychology, 14*, 1122200. <https://doi.org/10.3389/fpsyg.2023.1122200>
- Ashraf, H., Sodergren, M. H., Merali, N., Mylonas, G., Singh, H., & Darzi, A. (2018). Eye-tracking technology in medical education: A systematic review. *Medical Teacher, 40*(1), 62–69. <https://doi.org/10.1080/0142159X.2017.1391373>
- Behrend, J., & Dehais, F. (2020). How role assignment impacts decision-making in high-risk environments: Evidence from eye-tracking in aviation. *Safety Science, 127*, 104738. <https://doi.org/10.1016/j.ssci.2020.104738>
- Cak, S., Say, B., & Misirlisoy, M. (2020). Effects of working memory, attention, and expertise on pilots' situation awareness. *Cognition, Technology & Work, 22*(1), 85–94. <https://doi.org/10.1007/s10111-019-00551-w>
- Castro, C. (2008). *Human factors of visual and cognitive performance in driving*. CRC Press.
- Endsley, M. R. (1995). Measurement of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(1), 65–84. <https://doi.org/10.1518/001872095779049499>
- Endsley, M. R. (2015). Situation Awareness Misconceptions and Misunderstandings. *Journal of Cognitive Engineering and Decision Making, 9*(1), 4–32. <https://doi.org/10.1177/1555343415572631>
- Endsley, M. R. (2017). Toward a theory of situation awareness in dynamic systems. In *Situational awareness* (pp. 9–42).

- Endsley, M. R. (2018). Expertise and Situation Awareness. In *K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams. The Cambridge Handbook of Expertise and Expert Performance.* (pp. 714–742). Cambridge: Cambridge University Press.
- Endsley, M. R., & Garland, D. J. (2000). *Situation awareness analysis and measurement.* CRC press.
- Fang, R., Zhou, Q., Zhou, C., Yuan, S., Wang, K., Li, Q., Zhang, Y., & Li, J. (2025). A Systematic Review of Empirical Studies on Situation Awareness: Perspectives From the Interaction Among Humans, Machines, and the Task Environment. *PsyCh Journal*, 14(5), 718–733. <https://doi.org/10.1002/pchj.70027>
- Fetsch, C. R., Kiani, R., & Shadlen, M. N. (2014). Predicting the Accuracy of a Decision: A Neural Mechanism of Confidence. *Cold Spring Harbor Symposia on Quantitative Biology*, 79, 185–197. <https://doi.org/10.1101/sqb.2014.79.024893>
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00443>
- Gilbey, A., Walmsley, S., Tani, K., & Reweti, S. (2021). Decision making dyads and judgement overconfidence: Implications for high-risk industries. *Applied Ergonomics*, 97, 103529. <https://doi.org/10.1016/j.apergo.2021.103529>
- Grech, M., Horberry, T., & Koester, T. (2019). *Human factors in the maritime domain.* CRC press.
- Grech, M. R., Horberry, T., & Smith, A. (2002). Human Error in Maritime Operations: Analyses of Accident Reports Using the Leximancer Tool.

- Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(19), 1718–1721. <https://doi.org/10.1177/154193120204601906>
- Groeger, J. A. (2013). *Understanding Driving* (0 ed.). Routledge. <https://doi.org/10.4324/9780203769942>
- Horrey, W. J., Lesch, M. F., Mitsopoulos-Rubens, E., & Lee, J. D. (2015). Calibration of skill and judgment in driving: Development of a conceptual framework and the implications for road safety. *Accident Analysis & Prevention*, 76, 25–33. <https://doi.org/10.1016/j.aap.2014.12.017>
- Huang, H., Li, Z., Cheng, H., Wang, H., Jiang, J., Li, X., & Zgonnikov, A. (2025). Understanding driver cognition and decision-making behaviors in high-risk scenarios: A drift diffusion perspective. *Accident Analysis & Prevention*, 220, 108123. <https://doi.org/10.1016/j.aap.2025.108123>
- Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1322–1337. <https://doi.org/10.1098/rstb.2012.0037>
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126(4), 349–370. <https://doi.org/10.1037/0096-3445.126.4.349>
- Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816789.012>
- Koriat, A. (2015). Metacognition: Decision making Processes in Self-monitoring and Self-regulation. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell*

Handbook of Judgment and Decision Making (1st ed., pp. 356–379). Wiley.
<https://doi.org/10.1002/9781118468333.ch12>

Martinez-Marquez, D., Pingali, S., Panuwatwanich, K., Stewart, R. A., & Mohamed, S. (2021). Application of Eye Tracking Technology in Aviation, Maritime, and Construction Industries: A Systematic Review. *Sensors*, 21(13), 4289. <https://doi.org/10.3390/s21134289>

McAnally, K. I., Morris, A. P., & Best, C. (2017). Metacognitive monitoring and control in visual change detection: Implications for situation awareness and cognitive control. *PLOS ONE*, 12(9), e0176032. <https://doi.org/10.1371/journal.pone.0176032>

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502–517. <https://doi.org/10.1037/0033-295X.115.2.502>

Proctor, R. W., & Van Zandt, T. (2018). *Human factors in simple and complex systems*. CRC press.

Reale, C., Salwei, M. E., Militello, L. G., Weinger, M. B., Burden, A., Sushereba, C., Torsher, L. C., Andreae, M. H., Gaba, D. M., McIvor, W. R., Banerjee, A., Slagle, J., & Anders, S. (2023). Decision-Making During High-Risk Events: A Systematic Literature Review. *Journal of Cognitive Engineering and Decision Making*, 17(2), 188–212. <https://doi.org/10.1177/15553434221147415>

Salmon, P. M., Walker, G. H., & Stanton, N. A. (2016). Pilot error versus sociotechnical systems failure: A distributed situation awareness analysis of Air France 447. *Theoretical Issues in Ergonomics Science*, 17(1), 64–79. <https://doi.org/10.1080/1463922X.2015.1106618>

Salvendy, G. (2012). *Handbook of human factors and ergonomics*. (4th Ed.). John

Wiley & Sons.

Stanton, N. A., Chambers, P. R. G., & Piggott, J. (2001). Situational awareness and safety. *Safety Science*, 39(3), 189–204.
[https://doi.org/10.1016/S0925-7535\(01\)00010-8](https://doi.org/10.1016/S0925-7535(01)00010-8)

The newsletter of the IEA Issue 65—November 2000. (2000). *Ergonomics*, 43(11), 1939–1945. <https://doi.org/10.1080/00140130050174554>

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, 1(1), 3–14.
<https://doi.org/10.1007/s11409-006-6893-0>

Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). *Engineering psychology and human performance*. (5th ed.). Routledge.

Wise, J. A., Hopkin, V. D., & Garland, D. J. (2016). *Handbook of aviation human factors*. (2nd Ed.). CRC press.

Wohleber, R. W., & Matthews, G. (2016). Multiple facets of overconfidence: Implications for driving safety. *Transportation Research Part F: Traffic Psychology and Behaviour*, 43, 265–278.
<https://doi.org/10.1016/j.trf.2016.09.011>

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1310–1321.
<https://doi.org/10.1098/rstb.2011.0416>

Zhang, Z., Guo, B. H. W., Chang-Richards, A., Feng, Z., Jin, R., Zou, Y., & Miang Goh, Y. (2023). Digital technology enhanced situation awareness for

construction safety: Systematic review and future research directions. *Safety Science*, 167, 106280. <https://doi.org/10.1016/j.ssci.2023.106280>

2

Skill Transfer and Gaze Strategies Studied by Eye-Tracking: A Systematic Review

A study by

Lupia^a E., Bortolotti^a A., and Palumbo^a R. (2024)

^aDepartment of Neuroscience, Imaging and Clinical Sciences, University “G.D'Annunzio” of
Chieti-Pescara, Via Luigi Polacchi, 11, 66100 CH Chieti Scalo, Italy.

Abstract

Domain-specific expertise confers perceptual and cognitive benefits in engaging with information from that domain. Eye-tracking research has provided significant insight into the differences in performance between experts and novices across various fields. In visually demanding tasks, expert performance is characterized by systematic differences in gaze behavior relative to novice performance. The present systematic review examines empirical evidence on gaze-based skill transfer, to consolidate literature describing the application of expert gaze behavior as an adjunct tool to traditional instructional intervention. Following PRISMA guidelines, a structured search was conducted on Scopus database covering the period from 2000 to January 2024. Eight empirical studies met the inclusion criteria, four studies comparing expert and novice gaze behavior, and four studies evaluating the effects of gaze-based training interventions. Key findings demonstrate that, across domains, experts consistently exhibited more selective, anticipatory, and task-relevant visual attention patterns than novices. Also, gaze-trained individuals exhibit a performance advantage over traditionally trained counterparts. The consistency of these effects varied depending on task characteristics, and prior knowledge. In conclusion, the review offers practical insights on how gaze-based training can facilitate aspects of task performance and summarize the benefits of using expert gaze as a visual cue in terms of attention-modeling processes, human performance and safety.

2.1 Introduction

Expertise shapes how individuals perceive, elaborate, and keep in mind information from a specific task (Bédard and Chi, 1992). Over numerous fields and sensory modalities, domain-specific mastery confers perceptual and cognitive benefits (Ericsson, 2004). Within visual domains, these advantages can include the capacity to proficiently extract crucial data from the environment, rapidly perceive and process that data, and act accordingly (Papesh et al., 2021). Several theoretical accounts have clarified the processes that underlie expertise in the comprehension of visualizations. *Long-term working memory* theory (Ericsson & Kintsch, 1995) assumes that expertise extends the capacities for information processing thanks to the acquisition of recovery structures. The *Information-reduction* hypothesis (Haider & Frensch, 1999) suggests that expertise enhances efficiency by disregarding unnecessary information and concentrating on crucial task elements. Complementarily, the *holistic model of image perception* (Kundel et al., 2007) proposes that expertise expanded the visual span, allowing for more comprehensive image analysis.

A substantial body of research comparing experts and novices performance in visual search tasks has shown that, with growing experience, professional visual searchers needed to visit objects less frequently, they were less likely to view distractors, they viewed objects for shorter durations, and showed evidence of an expanding visual span (Wilson et al. 2010a; 2010b; O'Neill et al. 2011; Khan et al., 2012; Wood et al. 2013; Bond et al. 2014; Harvey et al. 2014; Jaarsma et al. 2014; Tien et al. 2015; Papesh et al., 2021; Reingold & Sheridan, 2023). These gaze patterns reflect strategic distribution of attention to task-critical information (Wilson et al., 2010a, 2010b), and reduced susceptibility to distractors (Litchfield & Donovan, 2016). The transition from novice to expert performance is also characterized by broader adaptations, including improvements in biomechanical and metabolic efficiency and increased neural efficiency, resulting in more selective and effective processing (Abernethy et al., 2007). From the perspective of learning psychology, this progression is in line with the model proposed by Fitts and Posner (1967), according to which performance shifts from declarative and tiring control to automated procedural execution with reduced attention demands.

Expert-novice comparisons provide a necessary basis for developing gaze-based training approaches, clarifying which perceptual and attentional mechanisms characterize skilled performance. Based on this framework, eye-tracking methodology has been used to support training interventions aimed at modeling novice gaze behavior in dynamic high-pressure environments (Yarrow et al., 2009; Wilson et al., 2010a; 2010b; Vine et al. 2012; Causer et al. 2014; Ashraf et al., 2017). A key motivation for such an approach lies in the possibility of accessing tacit expert knowledge, which is typically difficult to externalize through conventional instructional formats such as courses or written materials (Sik-wah Fong & Chu, 2006; Chetwood et al., 2012; Dzeng et al., 2016; Menezes et al., 2018). As a result, extracting expert knowledge is challenging (Mahroeian & Forozia, 2012), but eye-tracking technology offers a means to capture and transfer expert knowledge for improving trainees' performance in complex visual domains (Lohmeyer et al., 2013; van Meeuwen et al., 2014). In addition, empirical studies indicate that training gaze behavior can mediate the negative effects of anxiety on performance (Causer et al. 2014) and achieve faster completion times and fewer errors in performing a task (Vine et al., 2012).

2.1.1 Objectives

This systematic review brings together empirical work on expertise-related gaze behavior and on gaze-based interventions used to support learning and task performance. Previous literature reviews have addressed eye tracking within specific domains, most notably medical settings (Ashraf et al., 2018) and high-risk industries such as aviation, maritime, and construction (Martinez-Marquez et al., 2021), where the emphasis has largely been on assessment and feedback. Consideration of skill transfer has emerged more recently and has been discussed only in relation to a single training approach (e.g., eye movement modeling examples: Xie et al., 2021; Tunga & Cagiltay, 2023). Adopting a cross-domain perspective, the aims of this systematic review are:

- (1) To summarize visual organizational strategies that differentiate experts from novices;

- (2) To examine the training potential of eye movements as instructional cues, evaluating the feasibility and impact of transferring expert skills through gaze-based interventions across different tasks and approaches.

By integrating these lines of evidence, this review wants to clarify how gaze behavior associated with expertise can inform the design of training programs aimed at improving novice performance.

The article's structure is as follows. Section 2 outlines the conceptual background guiding the literature search and classification framework. Section 3 details the review methodology and provides a descriptive analysis of the included studies. Section 4 presents the literature review's results. Section 5 discusses the theoretical and practical contributions obtained from the literature analysis, the limitations and promising future research avenues. Finally, Section 6 concludes with the key points.

2.2 Conceptual Background

2.2.1 Skill Transfer Approaches

Literature review studies in cognitive science and learning (Xie et al., 2021; Tunga & Cagiltay, 2023) confirmed that the most common approach of gaze-based training is “Eye Movement Modeling Examples” (EMME), which focuses on recording the gaze behavior of experts while performing a task and presenting it to students in the form of video demonstrations. By observing expert gaze patterns, learners can infer which task elements are diagnostically relevant, how long they are inspected, and in what sequence information is processed (Jarodzka et al., 2012; Renkl, 2014). To provide a comprehensive learning experience, EMMEs are often complemented by verbal explanations detailing the expert's thought process, decision-making rationale, and interpretive strategies. By combining visual cues with verbal explanations, EMMEs create a multi-modal learning environment (Jarodzka et al., 2013; Van Gog et al., 2009). EMMEs have also been linked to improvements in metacognitive strategies, including the acquisition of expert heuristics, monitoring of task progress, and adaptive learning strategies (Collins & Kapur, 2006). Theoretically rooted in observational learning (Bandura, 1977), consistent with the Cognitive Theory of Multimedia Learning (Mayer, 2014) and the Cognitive Apprenticeship framework (Collins & Kapur, 2006), EMMEs

are expected to support learners' comprehension strategies by facilitating the selection, organization, and integration of task-relevant information (Groenendijk et al., 2013). From a neurocognitive perspective, the effectiveness of observational learning has also been discussed in relation to action observation mechanisms, including the mirror system (Bonini et al., 2022), which supports the internal simulation of observed actions and may facilitate the integration of new skills into the observer's behavioral repertoire (Calvo-Merino et al., 2005, 2006).

Another approach has been explored with the use of "Quiet Eye" Training (QET), targeting the modulation of the quiet eye (QE) period, which refers to the final fixation or tracking gaze preceding a critical motor action (Vickers, 1996, 2009; Klostermann, 2020). Longer QE periods have been linked with expert performance, particularly in tasks requiring precise motor control under time pressure, in sport, military operations and medical procedures (Causer et al., 2010; Ward et al., 2008). The QE is theorized to represent the time required to organize neural processes underlying action execution, with longer durations observed as task complexity increases (Williams et al., 2002). Also, maintaining a longer QE has also been shown to mitigate the detrimental effects of anxiety on performance, highlighting its role in attentional control under stress (Causer et al., 2011, 2014).

Several theoretical frameworks support the use of attention cueing and expert gaze observation for improving task performance. Attention cueing has been shown to promote selective processing of relevant information, reduce cognitive load, and enhance memory for cued elements (De Koning et al., 2009). Theory of Hints (Kirsh, 2009), posits that eye movements can function as implicit or explicit environmental aids, or "hints," that guide problem solving by shaping attention and interpretation (Thomas & Lleras, 2009a, 2009b; Ball & Litchfield, 2017). Observing expert gaze behavior may further support the formation of organized mental representations by revealing relational structures within visual scenes and highlighting task-relevant associations (Becchio et al., 2008; Jarodzka et al., 2013; Lobmaier et al., 2006). Taken together, these mechanisms suggest that expert eye movements can facilitate not only immediate performance improvements but also the development of transferable visual cognitive strategies applicable to novel situations (Kieras & Bovair, 1984).

2.2.2 Gaze-based Training Interventions: Application Fields

Medical Fields

In the medical field, gaze-based training was first introduced by Wilson and colleagues (2011), and has since been applied across several surgical and diagnostic contexts (Ashraf et al., 2018). Studies in laparoscopic training have revealed that novices exposed to expert gaze guidance develop more target-focused fixation patterns, resulting in faster task execution and fewer errors compared to discovery-learning or conventionally trained controls (Vine et al., 2012; Chetwood et al., 2012). In line with this evidence, Causer and colleagues (2014), demonstrated that while both traditional technical training and QET can improve surgical performance, QET uniquely supports performance stability under stress and anxiety, alongside more efficient gaze and motor coordination. Furthermore, neurophysiological findings suggest that combining visual guidance with expert gaze display may reduce cortical processing demands relative to verbal instruction alone, indicating more efficient visuocognitive processing during task execution (Leff et al., 2015). Gaze-based instructional formats have also been shown to benefit diagnostic learning. Jarodzka and colleagues (2012) showed that medical students' diagnostic accuracy improved most when expert narration was combined with gaze-based attention cueing, particularly when irrelevant visual information was suppressed using a spotlight technique. This finding highlights how important is the way gaze information is presented, rather than gaze display alone.

High-Risk Industries

In high-risk industries such as maritime operations, construction, and aviation, improving training programs is linked to the prevention of accidents and the enhancement of situational awareness (Martinez-Marquez et al., 2021).

In maritime applications, Li and colleagues (2020) proposed an innovative of gaze-based training approach, “expert-in-the-loop”, in which information about experts' gaze was used to decompose operational tasks, define critical areas of interest (AOIs), and generate weighted attention maps that reflect the visual priorities of experts. When applied in a simulated heavy lifting operation, trainees who received visual guidance derived from experts demonstrated more focused and safety-relevant attention patterns

than those who were only informed about operational risks, supporting the use of expert-based attention maps as training tools.

In the construction industry, eye tracking has been used to examine hazard recognition and visual search behavior. Pinheiro and colleagues (2016) demonstrated that simplified 2D representations led to more diffuse attention and faster but potentially superficial hazard identification compared to real images of buildings, suggesting limited ecological validity for complex environments. Extending this work, Jeelani and collaborators (2019) demonstrated that eye-tracking-based visual feedback derived from expert workers could improve novice hazard recognition performance. Specifically, broader and more systematic visual exploration of the workplace was associated with the identification of more hazards, highlighting the potential of personalized, gaze-based interventions for safety training.

In the aviation industry, eye tracking research has revealed differences in visual scanning strategies between different levels of expertise. Robinski and Stein (2013) found that experienced helicopter pilots relied more heavily on instrument information during critical phases of flight, while less experienced pilots showed greater dependence on external visual cues. Importantly, eye tracking feedback has been shown to support the transfer of skills from simulator training to real-world flight contexts, underscoring the value of gaze-based feedback in aviation training. In line with this evidence, previous work has shown that EMMEs can improve fault detection performance among novice aviation inspectors, helping to improve aviation safety (Sadasivan et al., 2005).

Sport Applications

Gaze-based training has been extensively examined in sport, most often through interventions based on the quiet eye (QE).

In golf putting, Moore and colleagues (2012) examined the effects of QET on performance as well as on kinematic (clubhead acceleration) and physiological responses (heart rate, muscle activity). Participants trained with the QE technique demonstrated superior accuracy, more effective gaze control, reduced clubhead acceleration, greater slowing of heart rate, and decreased muscle activity compared to those who received standard technical training. These findings suggest that quiet eye

training is associated with improved motor response preparation and a shift toward external focus.

Evidence from team sports points in a similar direction, although the picture is less uniform when pressure is introduced. In soccer penalty kicking, Wood and Wilson (2011) reported that QE-trained players showed improved gaze control and higher accuracy in retention tests. They also tested performance under pressure by simulating a penalty shootout, and during this high-stress scenario, the quiet eye-trained group maintained their improved visual focus, exhibiting longer-duration fixations on distant aiming points. Despite this, they were unable to sustain their accuracy advantage over the control group under these pressured conditions (Wood & Wilson, 2011).

In basketball, Vine and Wilson (2011) found that, when subjected to a high-pressure testing scenario, the quiet eye-trained athletes maintained their performance advantage. These findings lend credence to the potential of attentional training techniques in improving visuo-motor skills.

2.2.3 Eye-Tracking Metrics

Vision is the primary sensory system through which humans acquire information from the environment, guide action, and is involved in everyday interaction (Baritz & Lazar, 2019; Land & Hayhoe, 2001). During action, people often look at objects before manipulating them, which reflects ongoing cognitive processing and intentions (Hayhoe et al., 2003). Eye tracking research has identified several eye metrics that are linked to different cognitive, emotional, and physiological states, thereby providing a broader insight into the human mind (Skaramagkas et al., 2023). Eye-tracking studies commonly describe such behavior using measures like fixations, saccades, pupil size, and blink rate, usually analyzed within predefined Areas of Interest (Holmqvist et al., 2011). A summary of the most commonly used metrics in describing eye behaviors can be found in Supplementary Materials (Table A1).

Fixation

Fixation, a state where the eye and pupil remain stationary for about 180–300 ms, is typically linked with the acquisition of information and visual processing (Carter, &

Luke, 2020). During this period, individuals gather new data from a location, stimulus, or object. Various fixation metrics such as fixation frequency, fixation duration, maximum fixation duration, and the standard deviation of the fixation duration have been statistically analyzed and found to be associated with different cognitive attentional processes and human performance.

For instance, the distribution of fixations and fixation duration on relevant areas of interest (AOIs) can provide insights into a pilot's situation awareness (SA) performance and their level of expertise (Li et al., 2019). A study in aviation research discovered that seasoned pilots tend to distribute their fixations across multiple flight instruments, without lingering excessively on any single one. Interestingly, gaze behaviors can vary among experts from different fields. In the maritime industry, for example, expert operators often fixate more on the external environment, while novices tend to focus more on the deck. This variation can also be observed within the aviation industry itself, where commercial pilots often spend more time observing the instrument panel, while military pilots tend to look out the window more (Ziv, 2016). During a visual search activity, the duration of fixation can be used to predict an individual's ability to recognize hazards. On the other hand, a shorter fixation duration is generally associated with individuals who are experiencing anxiety or a state of threat (Vine et al., 2015).

Saccades

Saccades, which are quick eye movements that take place when an individual shifts their gaze between different points of fixation, typically last between 10 and 100 ms (Carter & Luke, 2020). During this brief period, the transfer of visual information is inhibited, indicating that saccades are not directly involved in cognitive processing. Nevertheless, the speed of saccades can be indicative of fatigue, stress, and lethargy (Schleicher et al., 2008).

For instance, the rate of saccades has been observed to decrease in response to fatigue and challenging tasks. The length of saccadic movements has been utilized as a measure of mental workload, and it has been found to increase in correlation with an

increase in mental workload (MWL) (Leigh & Kennard, 2004). Notably, very short saccades are associated with the presence of conflict.

Pupil Size

The pupil, one of the most notable features of the human eye, is easily observable and its size and center location can be readily determined using video recording (Kassner et al., 2014). The size of the pupil, which controls the amount of light that reaches the retina, is influenced by various factors including illumination, emotions, muscular fatigue, cognitive processes, and mental workload (MWL) (Lim et al., 2020).

For instance, an increase in pupillary dilation has been associated with a higher processing load in various maritime settings. Pupil size has also been utilized to understand the cognitive impact of stress during aviation emergencies and conflicts. In the context of quay crane operators, changes in pupil dilation have been used as a measure of alertness and the progression of fatigue (Murata, 1997).

Quiet Eye Movement

The “Quiet Eye” (QE) is a term used in the field of visual perception and motor performance. It refers to the final fixation or tracking gaze at a task-relevant location before the initiation of the final phase of a movement (Lebeau et al., 2016). This phenomenon is often observed in elite athletes and is believed to exemplify expertise-related differences in fixation behavior. The core concept behind Quiet Eye revolves around the gaze behavior that precedes an action.

For instance, before a basketball player releases a free throw, they often fixate their gaze on the basket, absorbing vital information before initiating the movement (Rienhoff et al., 2016). The “Quiet Eye” refers to this period of stable focus. The Quiet Eye technique can be applied by focusing on the object and/or target of your task. This can be achieved by keeping your eye movement steady and paying attention to a ball if you’re playing a game or the keys of a piano if you’re playing music. Research has shown that this technique can help athletes maintain their focus under high pressure, preventing them from ‘choking’ at crucial moments (Hill et al., 2010). It’s not just beneficial for athletes, but also for professionals like doctors performing keyhole

surgery, and it is of increasing interest to the military. In summary, the Quiet Eye technique is a fascinating area of research that offers insights into how visual focus can enhance performance in various fields, from sports to surgery.

2.3 Methodology

2.3.1 Literature Search and Selection Strategy

Systematic reviews are conducted with the purpose of identifying and summarizing existing evidence relating to a clearly defined research area, making the selection process transparent and replicable (Petticrew & Roberts, 2006).

In this review, a systematic search was performed to identify evidence-based studies investigating the transfer of skills through the use of information provided by expert observation. The literature search was carried out exclusively on the Scopus database on 05 February 2024, focusing on the period between 2000 to January 2024, and following the PRISMA guidelines (Page et al., 2021).

The literature search and selection strategy (Figure 2.1), consisted of three main steps: (a) defining relevant keywords, (b) searching one scholarly database and (c) identifying relevant papers. Three groups of keywords were established based on the conceptual framework described above. The first set of keywords is related to skill transfer (“skill” AND “transfer”); the second and third sets of keywords were selected based on the method through which skills transfer is achieved (“eye” AND “tracking”), (“gaze” AND “behavior”). These three groups of keywords were connected by Boolean operators in this sequence: (“skill” AND “transfer”) AND (“eye” AND “tracking”) OR (“gaze” AND “behavior”). Articles containing at least one relevant keyword in their title, abstract, or keyword list were selected for further examination.

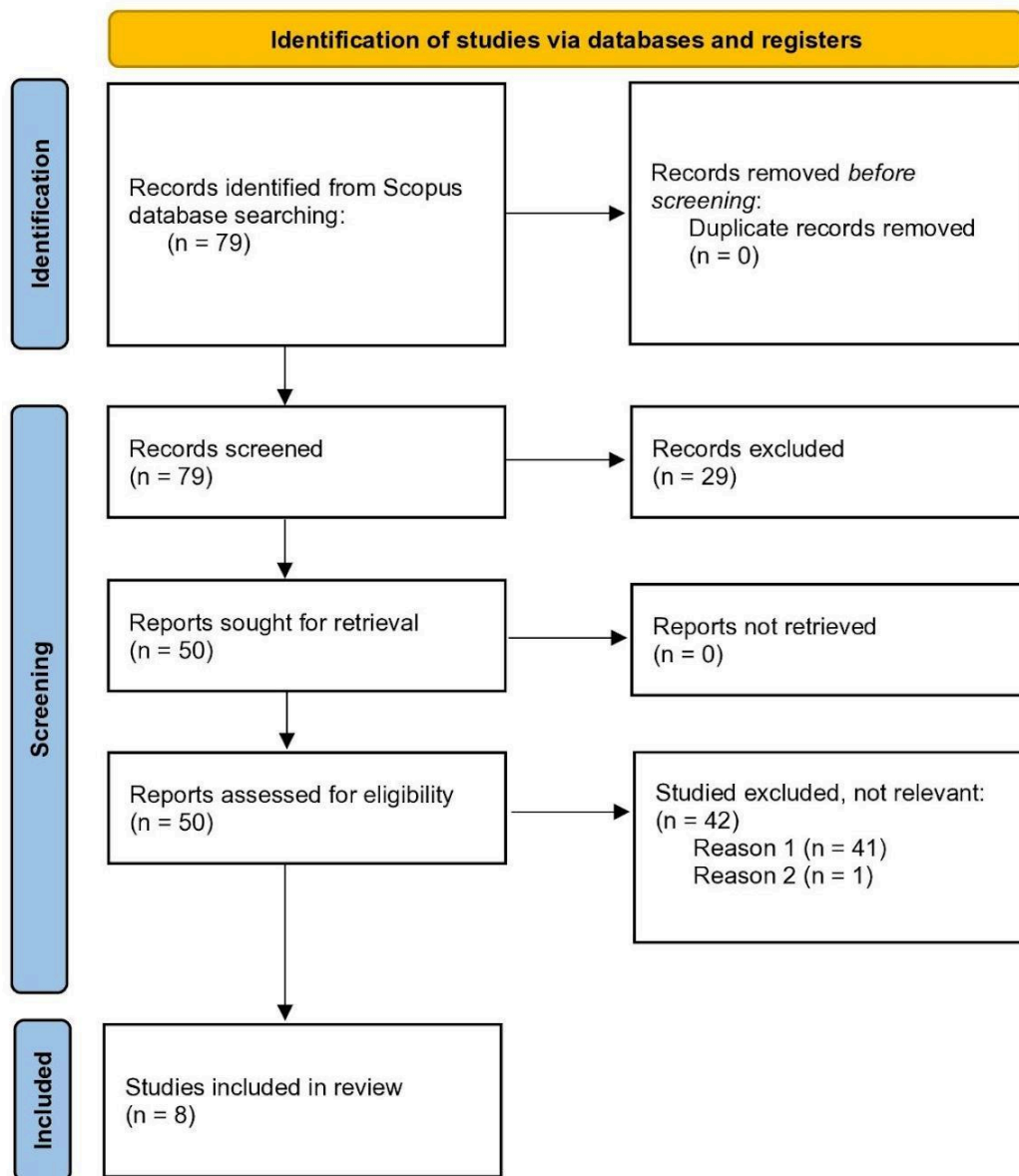


Figure 2.1 Literature selection flow diagram based on the PRISMA approach.

The initial search yielded 79 articles. An initial selection was made by reviewing titles, abstracts, keywords, and publication channels to assess their relevance and scope. At this stage, results that were review articles, conference proceedings, or book chapters, or that involved populations with atypical development, were excluded. Following this initial selection, 29 results were eliminated.

The remaining 50 records were organized in a shared Excel spreadsheet and evaluated according to predefined inclusion criteria. To be selected, studies had to investigate differences in visual behavior related to skills or the training potential of expert gaze applications in different fields. Articles were excluded if they did not report relevant empirical data (Reason 1) or were not published in English (Reason 2). All authors participated in this screening phase to reduce selection bias, and discrepancies were discussed until consensus was reached. The spreadsheet was used to document each decision during the filtering and screening process. After this step, the remaining articles were imported into Zotero © Reference Manager for full-text review. Studies were included in the final sample if they examined differences in eye movements as a function of experience or reported empirical findings on skill transfer through gaze-based modeling.

The selection process identified 8 studies suitable for final review.

2.3.2 Descriptive Analysis

This section provides a descriptive overview of the eight studies included in the review through a term co-occurrence analysis based on titles and abstracts. The analysis was conducted to offer a qualitative indication of the main concepts recurring across the selected papers rather than to support inferential claims.

Figure 2.2 illustrates the co-occurrence network generated using VOSviewer software. The clusters of terms identify sets of words that appear most often together in the titles and abstracts, using a binary counting method, so that repeated occurrences of a term within the same document were counted only once. A minimum occurrence threshold of two was applied, resulting in the selection of 14 terms from an initial set of 23. These terms formed a network consisting of 14 nodes organized into three clusters, represented by different colors in the visualization. The size of the nodes reflects the frequency of occurrence in the publications, while the thickness of the links represents the strength of the association between the terms.

The terms included were: skill (7), study (4), transfer (4), performance (3), eye movement (3), expert (3), time (3), learning (2), example of eye movement modeling (2), eye tracking (2), gaze (2), visual behavior (2), novice (2), and cardiovascular

intervention (2). Terms such as participant, model, EMME, and domain-specific acronyms (e.g. catheter, CBCVI) were excluded due to their low relevance or limited conceptual contribution.

The association strength method has been chosen for the analysis. The software identified three clusters: [1] The first cluster grouped terms related to the comparison of skills and the context of the activity (expert, beginner, gaze, visual behavior, time, transfer, cardiovascular intervention); [2] The second cluster focused on learning and performance outcomes (skill, study, learning, performance, eye movement, example of eye movement modeling); [3] The third cluster consisted only of the term “eye tracking”, reflecting its role as a methodological anchor across studies rather than a thematic focus.

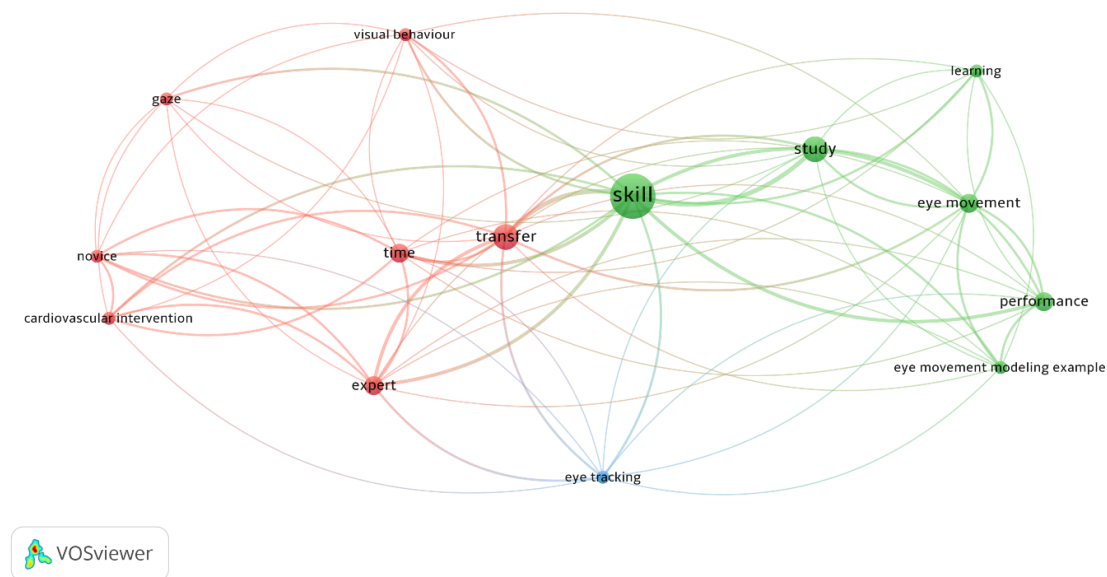


Figure 2.2 *VOSviewer visualization of a term co-occurrence network.*

2.4 Key findings

The systematic literature search identified eight studies that met the inclusion criteria. For analytical clarity, the studies were grouped into two categories of studies: (1) examining expertise-related differences in eye movement, and (2) evaluating the potential of gaze-based approaches for skill transfer.

2.4.1 Expertise-Related Differences in Perceptual Processing

Four of the included studies examined differences in eye movements between experienced and novice in medical settings. These studies focused primarily on surgical and interventional procedures and used eye tracking to describe how visual attention is distributed during task performance.

Zheng and colleagues (2011) compared experienced and novice surgeons during a simulated laparoscopic cholecystectomy in a virtual reality trainer while wearing a head-mounted eye tracker. Some patients were “preprogrammed” to have a mildly unstable cardiac condition during the procedure. During this task, the surgeons had to interact with the virtual environment to simulate the steps involved in a laparoscopic cholecystectomy procedure. Surgical performance, mental workload, and eye movements were recorded and compared between 13 experienced and 10 novice surgeons. Results revealed that although the experienced surgeons did not outperform the novices in terms of task completion time or error rates, differences in attention distribution emerged. Novices performed faster but focused more their attention on the surgical task, while experienced surgeons shifted their gaze more frequently to the anesthesia monitor to check vital signs. This pattern was observed in both stable and unstable patient conditions. In addition, experts reported less physical effort but higher levels of frustration, suggesting a different subjective experience of the demands of surgery. The authors interpreted these findings as evidence that vigilance during surgery is related to experience rather than speed or accuracy alone.

In cardiovascular interventions, Lanzer and authors (2017) aimed to analyze gaze behavior during real coronary angiography and catheter-based cardiovascular interventions (CBCVI). The researchers developed a computerized system to analyze eye-tracking data continuously across both time and space. This system processes information collected by a wearable eye-tracking device mounted on the user's head. Experienced operators exhibited stable viewing strategies, focusing predominantly on central regions of the screen and shifting their gaze only in response to relevant findings. In contrast, novices adopted more segmented and sequential scanning patterns. In particular, experts devoted a substantial portion of the procedure time to elements outside the screen, including patient monitoring, instrumentation, and communication

with the team. Although the study was limited to the operators' interaction with the X-ray screen, it provided evidence that expert performance is based on coordinated visual attention between task-relevant and contextual information sources.

In a similar way, Zimmermann and colleagues (2020) compared experienced and novice operators during real-life catheter-based cardiovascular interventions (CBCVIs) while wearing eye-tracking glasses. Experts showed fewer transitions between areas of interest and maintained a more focused gaze distribution during critical phases of the procedure, while novices showed higher transition rates and more diffuse visual attention. These findings reinforce the idea that expertise is reflected in more stable and selective visual strategies, which can be taken into account for training and assessment in interventional education.

Finally, Kulkarni and colleagues (2023) emphasized the importance of scene-dependent gaze metrics, that reflect both feedback and feedforward behaviors, to differentiate technical skill levels of trainees in laparoscopic surgery training. Using computer vision techniques, the authors identified gaze measures that distinguished multiple levels of competence rather than binary categories of experts and novices. Trainees with higher skill levels demonstrated anticipatory gaze behavior, characterized by looking ahead to elements of the upcoming task and reduced fixation on objects under direct manipulation. These feedforward gaze patterns were more sensitive to skill differences than traditional feedback-based metrics, suggesting their relevance for formative assessment and personalized training.

2.4.2 Effects of Gaze-based Interventions on Task Performance

Four studies examined the use of eye movements as an instructional signal to support task performance. Two studies were conducted in medical contexts, while two addressed non-medical tasks. Across these studies, gaze information was used to guide learners' attention toward task-relevant elements, either by training fixation behavior directly or by displaying expert gaze patterns as part of instructional examples.

In the medical setting, Wilson and colleagues (2011) conducted a randomized controlled trial to evaluate the impact of gaze training on the acquisition of laparoscopic skills under multitasking conditions. Thirty trainees with no laparoscopic experience

were assigned randomly to the gaze trained group, to the movement trained group, or discovery learning group. They performed the 'eye-hand coordination' task on a laparoscopic surgical simulator, where they had to touch flashing balls of different colors using specific instruments. Training consisted of ten trials of the eye-hand coordination task. They also concurrently performed a tone counting task to assess multitasking abilities. Gaze control was measured using a Mobile Eye gaze registration system. The results supported the hypotheses, because at the baseline all groups showed comparable performance and inefficient gaze strategies. After training, participants in the gaze training group completed the task more quickly and showed an increase in target-grasping fixations compared to the movement and discovery groups. Importantly, this performance advantage was maintained under multitasking conditions, despite no significant differences in the secondary tone-counting task. The authors attributed the observed benefits to changes in gaze behavior rather than improvements in instrument movement, suggesting that attention control played a central role in performance improvement. Gegenfurtner and colleagues (2017) examined the role of EMMEs in the interpretation of medical images. Twenty-three participants with varying levels of expertise observed an expert's gaze and verbal reasoning during the diagnosis of PET/CT images and were subsequently tested on retention and transfer. The study was structured in a progressive sequence of four distinct phases. In the baseline task participants diagnosed the first patient case, to establish their pre-existing diagnostic skills and knowledge. Then, in the learning phase subjects observed a modeling example of an expert model diagnosing a second patient case. In the retention task, participants were tasked with independently diagnosing the same case they had just seen the expert analyze. Finally, in the transfer task, participants were presented with a new patient case, for evaluating the participants' ability to transfer and apply the learned diagnostic processes to a novel situation. Results indicated that viewing the modeling example directed attention to diagnostically relevant regions. Performance and verbal reports suggested that experts benefited more from the intervention than novices, particularly in retention and transfer tasks. These findings indicate the role of prior knowledge in determining the effectiveness of gaze-based instruction in supporting adaptive skills.

Evidence supporting gaze-based teaching outside of medical training is mixed. Mason and colleagues (2016) used eye-movement modeling examples (EMMEs) in the school context, focusing on students' ability to integrate text and images when reading illustrated materials. Sixty-four students were randomly assigned to EMME or non-EMME conditions and were asked to read a text that was accompanied by relevant pictures. Students exposed to gaze reproductions showing a model's integrative reading strategy demonstrated more frequent gaze shifts between text and images and engaged in more rereading. In addition, the effectiveness of EMME varied based on students' reading comprehension abilities. Specifically, students with lower reading comprehension skills derived greater benefits from observing the strategic processing of text accompanied by illustrations. The authors concluded that EMME has the potential to enhance students' integrative processing of text and pictures, leading to improved learning outcomes and deeper comprehension of educational materials. In contrast, van Marlen and colleagues (2016) reported limited benefits of EMMEs in procedural problem-solving tasks through two experiments involving geometry problems. Results showed that viewing a model's eye movements did not lead to reliable improvements in accuracy or response time compared to standard solved examples. Although some trends suggested that meaningful visual cues could facilitate transfer, these effects were inconsistent and did not reach statistical significance, highlighting the sensitivity of EMME effectiveness to task complexity and instructional context.

In conclusion, gaze-based training shapes the allocation of visual attention and may improve task performance, but its effects vary across tasks and depend on task demands and prior experience.

2.5 Discussion

The present systematic review, adopting a cross-domain perspective, examined the expertise-related differences in gaze behavior during the comprehension of visualizations across different tasks, and whether gaze-based training interventions can improve task performance.

Expert-novice comparison of the studies included revealed that expert performance in a specific domain is characterized by more selective and stable visual

strategies. In particular, experts tend to allocate their attention among sources relevant to the task, keep track of contextual information, and show anticipatory gaze behavior rather than focusing only on the immediate action (Lanzer et al., 2017; Zimmermann et al., 2020). This pattern has been observed mainly in simulated and real medical procedures, such as laparoscopic surgery and cardiovascular catheter procedures (Zheng et al., 2011; Lanzer et al., 2017; Zimmermann et al., 2020). It is important to emphasize that these differences were not always accompanied by increased speed of task execution or reduced error rates (Zheng et al., 2011), suggesting that gaze behavior captures aspects of proficiency that are not fully reflected in conventional performance metrics. Moreover, the study conducted by Kulkarni and colleagues (2023) proves that anticipatory and feedforward gaze behaviors may be informative in differentiating between different levels of expertise. Eye tracking data provides access to process-level indicators of proficiency, such as vigilance, self-monitoring, and feedforward control (Sadasivan et al., 2005), which are particularly relevant in dynamic high-risk domains replete with stressors and distractions that may increase the attention demands (Fabri & Zayas-Castro, 2008; Szafranski et al., 2009; Poolton et al., 2011).

Concerning the effects of gaze-based interventions, the studies reviewed indicate that exposing trainees to expert gaze information can influence attention allocation and, in some cases, improve task performance. In the medical field, these interventions have been shown to promote the acquisition of technical skills and multitasking performance by promoting target-focused fixation patterns (Wilson et al., 2011). Skill transfer effects were also shown when gaze-related information was embedded in richer training contexts and supported higher-level cognitive processes, such as adaptive expertise in interpreting medical images (Gegenfurtner et al., 2017). This is particularly crucial in the medical field, where rapid advances in technology and evolving work practices require continually adapting established visual patterns to the new possibilities offered by the tasks at hand (Helle et al., 2011).

An additional finding seems to suggest that gaze-based modeling may, in some contexts, benefit highly skilled operators more than novices ones. Specifically, Gegenfurtner and authors (2017) reported that experienced medical professionals showed greater performance improvements than novices in interpreting unfamiliar

visualizations after observing the gaze behavior of experts. In this study, experts appeared to make more effective use of the information provided by gaze, particularly when dealing with cases that differed from their usual experience. This finding indicates that prior knowledge of the domain may influence how gaze-based guidance is used, especially in situations requiring adaptation to new working conditions, as discussed in relation to adaptive expertise (Hatano & Inagaki, 1986; Taya et al., 2013). At the same time, previous studies have shown that the transfer of skills to new contexts can be limited by cognitive biases developed through experience, which can restrict flexibility despite high levels of skill (Feltovich et al., 1997; Gegenfurtner & Seppänen, 2013). From this perspective, gaze-based interventions can support not only initial learning but also aspects of adaptive skills, helping experts to adjust established routines when faced with unfamiliar visual demands (Hatano & Inagaki, 1986).

Benefits have also been observed in school settings, where EMMEs have facilitated integrative processing of text and images, particularly for students with lower prior levels of expertise (Mason et al., 2016). These findings support the idea that an expert's gaze can serve as an implicit attention signal, guiding novices toward diagnostically relevant information and supporting cognitive and metacognitive processing.

Still, evidence for skill transfer through gaze-based training isn't consistent. Studies looking at procedural problem-solving tasks, such as geometry problems, have reported limited or inconsistent benefits from EMMEs (van Marlen et al., 2016). In fact, earlier research suggests that gaze-based modeling is more effective in non-procedural tasks that rely on visual inspection and classification (Jarodzka et al., 2012; Jarodzka et al., 2013). That is, for most procedural problem-solving tasks, the individual has to actively manipulate elements displayed on the screen (e.g., making calculations, selecting options, or moving objects). In line with the skill reversal effect, additional gaze information may impose unnecessary cognitive demands on more skilled students, potentially reducing its educational value (Kalyuga et al., 2003; Kalyuga, 2007). Overall, these results indicate that gaze-based training should not be considered a universally effective teaching tool. Rather, its impact seems to depend on several

interacting factors, including task complexity, the nature of visual information, the way visual stimuli are presented, and students' prior knowledge and skills.

Based on the included studies, what emerges is that the nature of domain-specific expertise is not merely reflected in faster or more accurate performance, but also in the way visual attention is organized, distributed, and anticipated during task performance. At the same time, evidence for the effectiveness of transferring these visual strategies through gaze-based training remains mixed and context-dependent.

2.5.1 Limitations and Implications

This work is subject to several limitations. First, the number of empirical studies directly addressing skill transfer through gaze-based interventions remains small, and methodological heterogeneity limits direct comparison between studies. In addition, because of the variability of the keywords used in the literature on this topic, we might have missed relevant papers. Second, eye tracking technology only captures overt visual attention (i.e., foveal gaze behavior, whereas visual attention in real-world activities also relies on covert mechanisms). Research has shown that attention can be allocated to peripheral or unfixed regions of the visual field, allowing individuals to process information without corresponding eye movements (Treisman & Gelade, 1980; Liversedge & Findlay, 2000). Consequently, eye tracking data may not fully reflect how visual information is interpreted, particularly in complex and dynamic environments where overt and covert attention mechanisms interact (Brunyé et al., 2019). This limitation implies that gaze measurements should be interpreted as partial indicators of visual cognition rather than direct proxies of attentional processing. A further limitation concerns the restricted set of potential moderators examined in the studies analyzed. Only a limited number of factors influencing the effectiveness of gaze-based interventions have been addressed systematically, leaving open some questions regarding the interaction between visual and verbal cues in examples of eye movement modeling and the variation of these effects between procedural and non-procedural tasks. Consequently, practitioners and educators should exercise caution when applying expert gaze models in educational settings and carefully consider task characteristics and student profiles.

Despite these limitations, the present review brings together existing evidence on expertise-related gaze behavior and on the use of gaze-based approaches in training, offering a consolidated view of a still fragmented literature. At the same time, it makes clear that several questions remain open, particularly concerning when gaze-based training is effective and for whom. Future work should therefore move beyond simple expert-novice comparisons and focus on how gaze behavior changes during learning, how it combines with other forms of instruction, and how it operates in more realistic task settings. Addressing these issues will likely require longitudinal studies and tasks with higher ecological validity in order to assess whether gaze-based interventions lead to learning effects that are stable and transferable.

2.6 Conclusion

The present work examined expertise-related differences in gaze behavior and the potential transfer of visual strategies through gaze-based training. The results suggest that, across studies and domains, expertise was linked to how visual attention was organized, monitored, or anticipated during task execution rather than in performance outcomes alone (such as reaction times or accuracy). Gaze-based training interventions were shown to influence the allocation of attention and, in some cases, task performance, particularly under demanding task conditions. However, their effectiveness varied across tasks and domains and appeared to depend on prior knowledge, and task demands.

References

- Abernethy, B., Maxwell, J. P., Masters, R. S. W., Kamp, J. V. D., & Jackson, R. C. (2007). Attentional Processes in Skill Learning and Expert Performance. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of Sport Psychology* (1st ed., pp. 245–263). Wiley. <https://doi.org/10.1002/9781118270011.ch11>
- Ashraf, H., Sodergren, M. H., Merali, N., Mylonas, G., Singh, H., & Darzi, A. (2018). Eye-tracking technology in medical education: A systematic review. *Medical Teacher*, 40(1), 62–69. <https://doi.org/10.1080/0142159X.2017.1391373>
- Ball, L. J., & Litchfield, D. (2017). Interactivity and Embodied Cues in Problem Solving, Learning and Insight: Further Contributions to a “Theory of Hints.” In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition Beyond the Brain* (pp. 115–132). Springer International Publishing. https://doi.org/10.1007/978-3-319-49115-8_6
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Baritz, M. I., & Lazar, A. M. (2019). Methodology for Monitoring the Behavior of the Visual System. *2019 E-Health and Bioengineering Conference (EHB)*, 1–4. <https://doi.org/10.1109/EHB47216.2019.8969931>
- Becchio, C., Sartori, L., Bulgheroni, M., & Castiello, U. (2008). Both your intention and mine are reflected in the kinematics of my reach-to-grasp movement. *Cognition*, 106(2), 894–912. <https://doi.org/10.1016/j.cognition.2007.05.004>
- Bédard, J., & Chi, M. T. H. (1992). Expertise. *Current Directions in Psychological Science*, 1(4), 135–139. <https://doi.org/10.1111/1467-8721.ep10769799>

- Bond, R. R., Zhu, T., Finlay, D. D., Drew, B., Kligfield, P. D., Guldenring, D., Breen, C., Gallagher, A. G., Daly, M. J., & Clifford, G. D. (2014). Assessing computerized eye tracking technology for gaining insight into expert interpretation of the 12-lead electrocardiogram: An objective quantitative approach. *Journal of Electrocardiology*, 47(6), 895–906. <https://doi.org/10.1016/j.jelectrocard.2014.07.011>
- Bonini, L., Rotunno, C., Arcuri, E., & Gallese, V. (2022). Mirror neurons 30 years later: Implications and applications. *Trends in Cognitive Sciences*, 26(9), 767–781. <https://doi.org/10.1016/j.tics.2022.06.003>
- Brunyé, T. T., Drew, T., Weaver, D. L., & Elmore, J. G. (2019). A review of eye tracking for understanding and improving diagnostic interpretation. *Cognitive Research: Principles and Implications*, 4(1), 7. <https://doi.org/10.1186/s41235-019-0159-2>
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action Observation and Acquired Motor Skills: An fMRI Study with Expert Dancers. *Cerebral Cortex*, 15(8), 1243–1249. <https://doi.org/10.1093/cercor/bhi007>
- Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or Doing? Influence of Visual and Motor Familiarity in Action Observation. *Current Biology*, 16(19), 1905–1910. <https://doi.org/10.1016/j.cub.2006.07.065>
- Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, 155, 49–62. <https://doi.org/10.1016/j.ijpsycho.2020.05.010>
- Causer, J., Bennett, S. J., Holmes, P. S., Janelle, C. M., & Williams, A. M. (2010). Quiet Eye Duration and Gun Motion in Elite Shotgun Shooting. *Medicine &*

- Science in Sports & Exercise*, 42(8), 1599–1608.
<https://doi.org/10.1249/MSS.0b013e3181d1b059>
- Causer, J., Holmes, P. S., Smith, N. C., & Williams, A. M. (2011). Anxiety, movement kinematics, and visual attention in elite-level performers. *Emotion*, 11(3), 595–602. <https://doi.org/10.1037/a0023225>
- Causer, J., Vickers, J. N., Snelgrove, R., Arsenault, G., & Harvey, A. (2014). Performing under pressure: Quiet eye training improves surgical knot-tying performance. *Surgery*, 156(5), 1089–1096.
<https://doi.org/10.1016/j.surg.2014.05.004>
- Chetwood, A. S. A., Kwok, K.-W., Sun, L.-W., Mylonas, G. P., Clark, J., Darzi, A., & Yang, G.-Z. (2012). Collaborative eye tracking: A potential training tool in laparoscopic surgery. *Surgical Endoscopy*, 26(7), 2003–2009.
<https://doi.org/10.1007/s00464-011-2143-x>
- Collins, A., & Kapur, M. (2006). *Cognitive apprenticeship (Vol. 291)*.
- De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a Framework for Attention Cueing in Instructional Animations: Guidelines for Research and Design. *Educational Psychology Review*, 21(2), 113–140.
<https://doi.org/10.1007/s10648-009-9098-7>
- Dzeng, R.-J., Lin, C.-T., & Fang, Y.-C. (2016). Using eye-tracker to compare search patterns between experienced and novice workers for site hazard identification. *Safety Science*, 82, 56–67.
<https://doi.org/10.1016/j.ssci.2015.08.008>
- Ericsson, K. A. (2004). Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains: *Academic Medicine*,

- 79(Supplement), S70–S81.
<https://doi.org/10.1097/00001888-200410001-00022>
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102(2), 211–245. <https://doi.org/10.1037/0033-295X.102.2.211>
- Fabri, P. J., & Zayas-Castro, J. L. (2008). Human error, not communication and systems, underlies surgical complications. *Surgery*, 144(4), 557–565. <https://doi.org/10.1016/j.surg.2008.06.011>
- Feltovich, P. J., Spiro, R. J., & Coulson, R. L. (1997). Issues of expert flexibility in contexts characterized by complexity and change. *Expertise in Context: Human and Machine.*, 125–146.
- Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Brooks/Cole.
- Gegenfurtner, A., Lehtinen, E., Jarodzka, H., & Säljö, R. (2017). Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis. *Computers & Education*, 113, 212–225. <https://doi.org/10.1016/j.compedu.2017.06.001>
- Gegenfurtner, A., & Seppänen, M. (2013). Transfer of expertise: An eye tracking and think aloud study using dynamic medical visualizations. *Computers & Education*, 63, 393–403. <https://doi.org/10.1016/j.compedu.2012.12.021>
- Groenendijk, T., Janssen, T., Rijlaarsdam, G., & Van Den Bergh, H. (2013). The effect of observational learning on students' performance, processes, and motivation in two creative domains. *British Journal of Educational Psychology*, 83(1), 3–28. <https://doi.org/10.1111/j.2044-8279.2011.02052.x>
- Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental*

- Psychology: Learning, Memory, and Cognition*, 25(1), 172–190.
<https://doi.org/10.1037/0278-7393.25.1.172>
- Harvey, A., Vickers, J. N., Snelgrove, R., Scott, M. F., & Morrison, S. (2014). Expert surgeon's quiet eye and slowing down: Expertise differences in performance and quiet eye duration during identification and dissection of the recurrent laryngeal nerve. *The American Journal of Surgery*, 207(2), 187–193.
<https://doi.org/10.1016/j.amjsurg.2013.07.033>
- Hatano, G., & Inagaki, K. (1986). Two courses of expertise. *Child Development and Education in Japan*, 262–272.
- Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, 3(1), 6.
<https://doi.org/10.1167/3.1.6>
- Helle, L., Nivala, M., Kronqvist, P., Gegenfurtner, A., Björk, P., & Säljö, R. (2011). Traditional microscopy instruction versus process-oriented virtual microscopy instruction: A naturalistic experiment with control group. *Diagnostic Pathology*, 6(S1), S8. <https://doi.org/10.1186/1746-1596-6-S1-S8>
- Hill, D. M., Hanton, S., Matthews, N., & Fleming, S. (2010). Choking in sport: A review. *International Review of Sport and Exercise Psychology*, 3(1), 24–39.
<https://doi.org/10.1080/17509840903301199>
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford university press.
- Jaarsma, T., Jarodzka, H., Nap, M., Van Merriënboer, J. J. G., & Boshuizen, H. P. A. (2014). Expertise under the microscope: Processing histopathological

- slides. *Medical Education*, 48(3), 292–300.
<https://doi.org/10.1111/medu.12385>
- Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., & Eika, B. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, 40(5), 813–827.
<https://doi.org/10.1007/s11251-012-9218-5>
- Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a Model's eye movements fosters learning. *Learning and Instruction*, 25, 62–70.
<https://doi.org/10.1016/j.learninstruc.2012.11.004>
- Jeelani, I., Albert, A., Han, K., & Azevedo, R. (2019). Are Visual Search Patterns Predictive of Hazard Recognition Performance? Empirical Investigation Using Eye-Tracking Technology. *Journal of Construction Engineering and Management*, 145(1), 04018115.
[https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001589](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001589)
- Kalyuga, S. (2007). Expertise Reversal Effect and Its Implications for Learner-Tailored Instruction. *Educational Psychology Review*, 19(4), 509–539.
<https://doi.org/10.1007/s10648-007-9054-3>
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist*, 38(1), 23–31.
https://doi.org/10.1207/S15326985EP3801_4
- Kassner, M., Patera, W., & Bulling, A. (2014). Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 1151–1160.
<https://doi.org/10.1145/2638728.2641695>

- Khan, R. S. A., Tien, G., Atkins, M. S., Zheng, B., Panton, O. N. M., & Meneghetti, A. T. (2012). Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation? *Surgical Endoscopy*, 26(12), 3536–3540. <https://doi.org/10.1007/s00464-012-2400-7>
- Kieras, D., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science*, 8(3), 255–273. [https://doi.org/10.1016/S0364-0213\(84\)80003-8](https://doi.org/10.1016/S0364-0213(84)80003-8)
- Kirsh, D. (2009). *Problem solving and situated cognition*. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition*. Cambridge: Cambridge University Press.
- Klostermann, A. (2020). Perception and action in a far-aiming task: Inhibition demands and the functionality of the Quiet Eye in motor performance. *Psychology of Sport and Exercise*, 50, 101736. <https://doi.org/10.1016/j.psychsport.2020.101736>
- Kulkarni, C. S., Deng, S., Wang, T., Hartman-Kenzler, J., Barnes, L. E., Parker, S. H., Safford, S. D., & Lau, N. (2023). Scene-dependent, feedforward eye gaze metrics can differentiate technical skill levels of trainees in laparoscopic surgery. *Surgical Endoscopy*, 37(2), 1569–1580. <https://doi.org/10.1007/s00464-022-09582-3>
- Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic Component of Image Perception in Mammogram Interpretation: Gaze-tracking Study. *Radiology*, 242(2), 396–402. <https://doi.org/10.1148/radiol.2422051997>
- Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25–26), 3559–3565. [https://doi.org/10.1016/S0042-6989\(01\)00102-X](https://doi.org/10.1016/S0042-6989(01)00102-X)

- Lanzer, P., Al-Naser, M., Bukhari, S. S., Dengel, A., & Krupinski, E. A. (2017). Eye tracking in catheter-based cardiovascular interventions: Early results. *Journal of Medical Imaging*, 4(3), 035502. <https://doi.org/10.1117/1.JMI.4.3.035502>
- Lebeau, J.-C., Liu, S., Sáenz-Moncaleano, C., Sanduvete-Chaves, S., Chacón-Moscoso, S., Becker, B. J., & Tenenbaum, G. (2016). Quiet Eye and Performance in Sport: A Meta-Analysis. *Journal of Sport and Exercise Psychology*, 38(5), 441–457. <https://doi.org/10.1123/jsep.2015-0123>
- Leff, D. R., James, D. R. C., Orihuela-Espina, F., Kwok, K.-W., Sun, L. W., Mylonas, G., Athanasiou, T., Darzi, A. W., & Yang, G.-Z. (2015). The impact of expert visual guidance on trainee visual search strategy, visual attention and motor skills. *Frontiers in Human Neuroscience*, 9. <https://doi.org/10.3389/fnhum.2015.00526>
- Leigh, R. J., & Kennard, C. (2004). Using saccades as a research tool in the clinical neurosciences. *Brain*, 127(3), 460–477. <https://doi.org/10.1093/brain/awh035>
- Li, G., Mao, R., Hildre, H. P., & Zhang, H. (2020). Visual Attention Assessment for Expert-in-the-Loop Training in a Maritime Operation Simulator. *IEEE Transactions on Industrial Informatics*, 16(1), 522–531. <https://doi.org/10.1109/TII.2019.2945361>
- Lim, J. Z., Mountstephens, J., & Teo, J. (2020). Emotion Recognition Using Eye-Tracking: Taxonomy, Review and Current Challenges. *Sensors*, 20(8), 2384. <https://doi.org/10.3390/s20082384>
- Litchfield, D., & Donovan, T. (2016). Worth a quick look? Initial scene previews can guide eye movements as a function of domain-specific expertise but can also have unforeseen costs. *Journal of Experimental Psychology: Human*

- Perception and Performance*, 42(7), 982–994.
<https://doi.org/10.1037/xhp0000202>
- Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4(1), 6–14.
[https://doi.org/10.1016/S1364-6613\(99\)01418-7](https://doi.org/10.1016/S1364-6613(99)01418-7)
- Lobmaier, J. S., Fischer, M. H., & Schwaninger, A. (2006). Objects Capture Perceived Gaze Direction. *Experimental Psychology*, 53(2), 117–122.
<https://doi.org/10.1027/1618-3169.53.2.117>
- Lohmeyer, Q., Meboldt, M., & Matthiesen, S. (2013). *Analysing visual strategies of novice and experienced designers by eye tracking application*.
- Mahroeian, H., & Forozia, A. (2012). *Challenges in Managing Tacit Knowledge: A Study on Difficulties in Diffusion of Tacit Knowledge in Organizations*.
<https://api.semanticscholar.org/CorpusID:18992452>
- Martinez-Marquez, D., Pingali, S., Panuwatwanich, K., Stewart, R. A., & Mohamed, S. (2021). Application of Eye Tracking Technology in Aviation, Maritime, and Construction Industries: A Systematic Review. *Sensors*, 21(13), 4289. <https://doi.org/10.3390/s21134289>
- Mason, L., Pluchino, P., & Tornatora, M. C. (2016). Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology*, 47(6), 1083–1095.
<https://doi.org/10.1111/bjet.12271>
- Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (1st ed., pp. 31–48). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816819.004>

- Menezes, P., Francisco, J., & Patrão, B. (2018). The Importance of Eye-Tracking Analysis in Immersive Learning—A Low Cost Solution. In M. E. Auer & D. G. Zutin (Eds.), *Online Engineering & Internet of Things* (Vol. 22, pp. 689–697). Springer International Publishing.
https://doi.org/10.1007/978-3-319-64352-6_65
- Moore, L. J., Vine, S. J., Cooke, A., Ring, C., & Wilson, M. R. (2012). Quiet eye training expedites motor learning and aids performance under heightened anxiety: The roles of response programming and external attention. *Psychophysiology*, 49(7), 1005–1015.
<https://doi.org/10.1111/j.1469-8986.2012.01379.x>
- Murata, A. (1997). Assessment of Fatigue by Pupillary Response. *IEICE Transactions on Fundamentals*, E80-A(7), 1318–1323.
https://globals.ieice.org/en_transactions/fundamentals/10.1587/e80-a_7_1318/#
- O'Neill, E. C., Kong, Y. X. G., Connell, P. P., Ong, D. N., Haymes, S. A., Coote, M. A., & Crowston, J. G. (2011). Gaze Behavior among Experts and Trainees during Optic Disc Examination: Does How We Look Affect What We See? *Investigative Ophthalmology & Visual Science*, 52(7), 3976.
<https://doi.org/10.1167/iovs.10-6912>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71.
<https://doi.org/10.1136/bmj.n71>
- Papesh, M. H., Hout, M. C., Guevara Pinto, J. D., Robbins, A., & Lopez, A. (2021). Eye movements reflect expertise development in hybrid search. *Cognitive*

- Research: Principles and Implications*, 6(1), 7.
<https://doi.org/10.1186/s41235-020-00269-8>
- Petticrew, M., & Roberts, H. (2006). *Systematic Reviews in the Social Sciences: A Practical Guide* (1st ed.). Wiley. <https://doi.org/10.1002/9780470754887>
- Pinheiro, R. B. D. O., Pradhananga, N., Jianu, R., & Orabi, W. (2016, July 21). *Eye-Tracking Technology for Construction Safety: A Feasibility Study*. 33th International Symposium on Automation and Robotics in Construction. <https://doi.org/10.22260/ISARC2016/0035>
- Poolton, J. M., Wilson, M. R., Malhotra, N., Ngo, K., & Masters, R. S. W. (2011). A comparison of evaluation, time pressure, and multitasking as stressors of psychomotor operative performance. *Surgery*, 149(6), 776–782. <https://doi.org/10.1016/j.surg.2010.12.005>
- Reingold, E. M., & Sheridan, H. (2023). *Chess expertise reflects domain-specific perceptual processing: Evidence from eye movements*.
- Renkl, A. (2014). Toward an Instructionally Oriented Theory of Example-Based Learning. *Cognitive Science*, 38(1), 1–37. <https://doi.org/10.1111/cogs.12086>
- Rienhoff, R., Tirp, J., Strauß, B., Baker, J., & Schorer, J. (2016). The ‘Quiet Eye’ and Motor Performance: A Systematic Review Based on Newell’s Constraints-Led Model. *Sports Medicine*, 46(4), 589–603. <https://doi.org/10.1007/s40279-015-0442-4>
- Robinski, M., & Stein, M. (2013). Tracking Visual Scanning Techniques in Training Simulation for Helicopter Landing. *Journal of Eye Movement Research*, 6(2). <https://doi.org/10.16910/jemr.6.2.3>

- Sadasivan, S., Greenstein, J. S., Gramopadhye, A. K., & Duchowski, A. T. (2005). Use of eye movements as feedforward training for a synthetic aircraft inspection task. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 141–149. <https://doi.org/10.1145/1054972.1054993>
- Schleicher, R., Galley, N., Briest, S., & Galley, L. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: Looking tired? *Ergonomics*, 51(7), 982–1010. <https://doi.org/10.1080/00140130701817062>
- Sik-wah Fong, P., & Chu, L. (2006). Exploratory Study of Knowledge Sharing in Contracting Companies: A Sociotechnical Perspective. *Journal of Construction Engineering and Management*, 132(9), 928–939. [https://doi.org/10.1061/\(ASCE\)0733-9364\(2006\)132:9\(928\)](https://doi.org/10.1061/(ASCE)0733-9364(2006)132:9(928))
- Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N., Tripoliti, E., Marias, K., Fotiadis, D. I., & Tsiknakis, M. (2023). Review of Eye Tracking Metrics Involved in Emotional and Cognitive Processes. *IEEE Reviews in Biomedical Engineering*, 16, 260–277. <https://doi.org/10.1109/RBME.2021.3066072>
- Szafranski, C., Kahol, K., Ghaemmaghami, V., Smith, M., & Ferrara, J. J. (2009). Distractions and surgical proficiency: An educational perspective. *The American Journal of Surgery*, 198(6), 804–810. <https://doi.org/10.1016/j.amjsurg.2009.04.027>
- Taya, S., Windridge, D., & Osman, M. (2013). Trained Eyes: Experience Promotes Adaptive Gaze Control in Dynamic and Uncertain Visual Environments. *PLoS ONE*, 8(8), e71371. <https://doi.org/10.1371/journal.pone.0071371>
- Thomas, L. E., & Lleras, A. (2009a). Covert shifts of attention function as an implicit aid to insight. *Cognition*, 111(2), 168–174. <https://doi.org/10.1016/j.cognition.2009.01.005>

- Thomas, L. E., & Lleras, A. (2009b). Swinging into thought: Directed movement guides insight in problem solving. *Psychonomic Bulletin & Review*, 16(4), 719–723. <https://doi.org/10.3758/PBR.16.4.719>
- Tien, T., Pucher, P. H., Sodergren, M. H., Sriskandarajah, K., Yang, G.-Z., & Darzi, A. (2015). Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. *Surgical Endoscopy*, 29(2), 405–413. <https://doi.org/10.1007/s00464-014-3683-7>
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Tunga, Y., & Cagiltay, K. (2023). Looking through the model's eye: A systematic review of eye movement modeling example studies. *Education and Information Technologies*, 28(8), 9607–9633. <https://doi.org/10.1007/s10639-022-11569-5>
- Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25(3), 785–791. <https://doi.org/10.1016/j.chb.2009.02.007>
- Van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2016). Showing a model's eye movements in examples does not improve learning of problem-solving tasks. *Computers in Human Behavior*, 65, 448–459. <https://doi.org/10.1016/j.chb.2016.08.041>
- Van Meeuwen, L. W., Jarodzka, H., Brand-Gruwel, S., Kirschner, P. A., De Bock, J. J. P. R., & Van Merriënboer, J. J. G. (2014). Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction*, 32, 10–21. <https://doi.org/10.1016/j.learninstruc.2014.01.004>

- Vickers, J. N. (1996). Control of Visual Attention during the Basketball Free Throw. *The American Journal of Sports Medicine*, 24(6_suppl), S93–S97. <https://doi.org/10.1177/036354659602406S25>
- Vickers, J. N. (2009). Advances in coupling perception and action: The quiet eye as a bidirectional link between gaze, attention, and action. In *Progress in Brain Research* (Vol. 174, pp. 279–288). Elsevier. [https://doi.org/10.1016/S0079-6123\(09\)01322-3](https://doi.org/10.1016/S0079-6123(09)01322-3)
- Vine, S. J., Masters, R. S. W., McGrath, J. S., Bright, E., & Wilson, M. R. (2012). Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills. *Surgery*, 152(1), 32–40. <https://doi.org/10.1016/j.surg.2012.02.002>
- Vine, S. J., Uiga, L., Lavric, A., Moore, L. J., Tsaneva-Atanasova, K., & Wilson, M. R. (2015). Individual reactions to stress predict performance during a critical aviation incident. *Anxiety, Stress, & Coping*, 28(4), 467–477. <https://doi.org/10.1080/10615806.2014.986722>
- Vine, S. J., & Wilson, M. R. (2011). The influence of quiet eye training and pressure on attention and visuo-motor control. *Acta Psychologica*, 136(3), 340–346. <https://doi.org/10.1016/j.actpsy.2010.12.008>
- Ward, P., Farrow, D., Harris, K. R., Williams, A. M., Eccles, D. W., & Ericsson, K. A. (2008). Training Perceptual-Cognitive Skills: Can Sport Psychology Research Inform Military Decision Training? *Military Psychology*, 20(sup1), S71–S102. <https://doi.org/10.1080/08995600701804814>
- VOSviewer (2023). VOSviewer [Computer software]. Centre for Science and Technology Studies, Leiden University, The Netherlands. <https://www.vosviewer.com/> (First public release 2009)

- Williams, A. M., Singer, R. N., & Frehlich, S. G. (2002). Quiet Eye Duration, Expertise, and Task Complexity in Near and Far Aiming Tasks. *Journal of Motor Behavior*, 34(2), 197–207. <https://doi.org/10.1080/00222890209601941>
- Wilson, M., Coleman, M., & McGrath, J. (2010a). Developing basic hand-eye coordination skills for laparoscopic surgery using gaze training. *BJU International*, 105(10), 1356–1358. <https://doi.org/10.1111/j.1464-410X.2010.09315.x>
- Wilson, M., McGrath, J., Vine, S., Brewer, J., Defriend, D., & Masters, R. (2010b). Psychomotor control in a virtual laparoscopic surgery training environment: Gaze control parameters differentiate novices from experts. *Surgical Endoscopy*, 24(10), 2458–2464. <https://doi.org/10.1007/s00464-010-0986-1>
- Wilson, M. R., Vine, S. J., Bright, E., Masters, R. S. W., Defriend, D., & McGrath, J. S. (2011). Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: A randomized, controlled study. *Surgical Endoscopy*, 25(12), 3731–3739. <https://doi.org/10.1007/s00464-011-1802-2>
- Wood, G., Knapp, K. M., Rock, B., Cousens, C., Roobottom, C., & Wilson, M. R. (2013). Visual expertise in detecting and diagnosing skeletal fractures. *Skeletal Radiology*, 42(2), 165–172. <https://doi.org/10.1007/s00256-012-1503-5>
- Wood, G., & Wilson, M. R. (2011). Quiet-eye training for soccer penalty kicks. *Cognitive Processing*, 12(3), 257–266. <https://doi.org/10.1007/s10339-011-0393-0>
- Xie, H., Zhao, T., Deng, S., Peng, J., Wang, F., & Zhou, Z. (2021). Using eye movement modelling examples to guide visual attention and foster cognitive performance: A meta-analysis. *Journal of Computer Assisted Learning*, 37(4), 1194–1206. <https://doi.org/10.1111/jcal.12568>

- Yarrow, K., Brown, P., & Krakauer, J. W. (2009). Inside the brain of an elite athlete: The neural processes that support high achievement in sports. *Nature Reviews Neuroscience*, 10(8), 585–596. <https://doi.org/10.1038/nrn2672>
- Zheng, B., Tien, G., Atkins, S. M., Swindells, C., Tanin, H., Meneghetti, A., Qayumi, K. A., & Panton, O. N. M. (2011). Surgeon's vigilance in the operating room. *The American Journal of Surgery*, 201(5), 673–677. <https://doi.org/10.1016/j.amjsurg.2011.01.016>
- Zimmermann, J. M., Vicentini, L., Lohmeyer, Q., Taramasso, M., Maisano, F., & Meboldt, M. (2020). Visual Behaviour Strategies of Operators during Catheter-Based Cardiovascular Interventions. *Journal of Medical Systems*, 44(1), 12. <https://doi.org/10.1007/s10916-019-1480-5>
- Ziv, G. (2016). Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation. *The International Journal of Aviation Psychology*, 26(3–4), 75–104. <https://doi.org/10.1080/10508414.2017.1313096>
- Zotero. (2020). Zotero [Computer software]. Corporation for Digital Scholarship. <https://Zotero.org> (Originally published 2006)

3

The Impact of Overconfidence and Environmental Conditions on Hazard Perception and Risk Assessment: An Experimental Study Using Video-Based Traffic Scenarios

A study by

Lupia^a E., Bortolotti^a A., and Palumbo^a R. (2025)

^aDepartment of Neuroscience, Imaging and Clinical Sciences, University “G.D'Annunzio” of Chieti-Pescara, Via Luigi Polacchi, 11, 66100 CH Chieti Scalo, Italy.

Abstract

Human decisions are accompanied by internal confidence judgments about the likelihood of being correct. These judgments are not always well-calibrated. Miscalibration occurs when there is a discrepancy between self-reported confidence regarding the correctness of performance and the objective performance. As a result, cognitive bias such as overconfidence can emerge, shaping how individuals perceive and behave. Traffic represents a distinctive decision environment, in which individuals have to constantly monitor and interpret changing perceptual information. Although overconfidence has been linked to unsafe driving, prior research conceptualizes confidence through self-report questionnaires and group-based analyses. Using performance-based measures, the present study aims to capture confidence as a continuous variable, investigating whether this bias can impact both the hazard perception performance and subjective experience of risk. Sixty-five participants completed a novel adaptation of the Hazard Perception Task (HPT), which integrated trial-by-trial confidence judgments and risk estimation of driving scenarios, alongside scores from the Hazard Perception Questionnaire (HPQ) and the Environmental Sensitivity (HSP scale). Results showed that overconfidence significantly reduced hazard perception accuracy, whereas reaction times to hazard and risk estimation were unaffected. Response accuracy declined also during nighttime in safe scenarios but remained high in hazardous trials. Subjective risk estimation was driven by the presence of hazards and lightning conditions. Environmental Sensitivity (HSP scale) showed a significant positive correlation with risk estimation. No relationship emerged between overconfidence and driving experience. Understanding how confidence judgments and individual differences operate in this high-risk context is therefore critical for safety.

3.1 Introduction

Metacognition, or "thinking about thinking", is a higher-level cognitive process that involves regulating and assessing one's mental operations, which is essential for self-regulation and learning (Rhodes, 2019; Norman et al., 2019; Efklides & Misailidi, 2010a; Efklides & Misailidi, 2010b). Confidence judgments, a key aspect of metacognition, refer to the ability to assess the quality of one's performance (Metcalf & Shimamura, 1994). This mechanism enables the evaluation of the correctness of one's own decisions and actions (Kepecs et al., 2012; Yeung & Summerfield, 2012; Fetsch et al., 2014). These judgments are not always accurate, and their miscalibration can lead to cognitive biases (Horrey et al., 2015). A relevant bias is overconfidence, which is commonly characterized by three distinct manifestations. Overestimation reflects the belief that one's abilities or success probability exceed objective measures. Overplacement, or the Above-Average Effect, refers to individuals that see themselves as more skilled than peers, and overprecision, an excessive confidence in the accuracy of one's beliefs (for a comprehensive review, see Moore & Healy, 2008).

Research demonstrates that confidence in a decision can increase with the accumulation of evidence supporting that decision, together with the decrease of evidence for an alternative one (Yeung & Summerfield, 2012; Kiani et al., 2014; Murphy et al., 2015), and overconfidence often linked to confirmation bias. Metacognitive calibration (e.g. the ability to properly judge one's decision or task performance; Horrey et al., 2015) improves when individuals consider disconfirming evidence (Koriat et al., 1980). These findings are well-supported in laboratory settings (Kiani et al., 2014; Chua & Solinger, 2015; Handel et al., 2020; Brus et al., 2021; for a review see Fleming, 2024). Confidence judgments are crucial in real-life settings, where sequential decisions must be made without immediate feedback (Fetsch et al., 2014). Within this context, overconfidence has serious implications in various professional fields, such as clinical judgment (Croskerry & Norman, 2008), investment decisions (Kumar & Prince, 2023), driving safety (Wohleber & Matthews, 2016), finance and trading (Chuang & Lee, 2006; Baker & Wurgler, 2013).

The mobility domain is a complex adaptive system in which human decisions occur in the presence of considerable dynamism and variability (Groeger, 2013). Individuals have to rapidly process, interpret and adapt their decision to changing conditions in real time (Hills, 1980; Warren, 2006; Owsley & McGwin, 2010), and simultaneously they need to evaluate their performance while maintaining a calibrated judgment of correctness of one's own actions (Horrey et al., 2015). Based on Deery's model (1999), young drivers are inclined to overestimate their driving abilities, while underestimating their risk of crash involvement. This tendency can lead individuals to believe they are less likely to experience negative outcomes than others (Matthews & Moran, 1986; Gregersen, 1996; Deery, 1999; Sümer et al., 2006; McKenna & Horswill, 2006). According to Horrey and colleagues (2015), overestimation reflects poor calibration, in which perceived and actual ability or performance diverge. In high-risk environments such as traffic, this miscalibration contributes to safety concerns by promoting increased risk-taking behaviors (Horrey et al., 2015), biased decision-making, and a distorted belief in one's capacity to control driving impairments, such as fatigue and distraction (Svenson, 1981; Matthews & Moran, 1986; Horswill & McKenna, 1999; Deery, 1999; Wohleber & Matthews, 2016).

This manuscript primarily aims to investigate the relationship between confidence bias and perception in driving. A performance-based approach was adopted to explore whether this bias affects both hazard perception and the subjective experience of risk. To this end, we developed a novel adaptation of the video-based Hazard Perception Task (HPT), integrating trial-by-trial confidence judgments and subjective risk estimation of road scenarios. Also, it investigates how individual differences in Environmental Sensitivity relate to subjective risk estimation of potential threat in driving, and finally examines the relationship between overconfidence and driving experience.

3.1.1 The role of Confidence in Hazard Perception

Driving involves a three-stage cognitive process: perceiving driving-related information, decision-making, and vehicle control actions (Groeger, 2013). Hazard Perception Skill (HPS) represents a critical component of the perception stage (Cao et al., 2022). In line with the definition proposed by Crundall and colleagues (2003),

Hazard Perception Skill is defined as the “driver’s ability to detect and respond in time and appropriately to potentially dangerous events on the road” (Åbele et al., 2018). HPS is typically assessed through behavioral task, where participants observe dynamic traffic scenes from driver-perspective, and a response to the hazard is requested (e.g. Wetton et al., 2011; Zeuwts et al., 2017; Sun et al., 2021; Crundall et al., 2021). The behavioral measures include reaction time to hazards, response accuracy in detecting road hazards, and eye movements (for review, see Moran et al., 2019; Cao et al., 2022). Poor HPS is associated with increased rates of crash involvement and unsafe driving behavior (McKnight & McKnight, 2003; Horswill & McKenna, 2004; Fisher et al., 2006; Pollatsek et al., 2006; Cheng et al., 2011; Horswill et al., 2015). Hazard perception ability can be affected by several cognitive variables (Cao et al., 2022), as well as environmental factors such as nighttime visibility and weather conditions (Garay et al., 2004; Konstantopoulos et al., 2010; Asadamraji et al., 2019; Evans et al., 2020; Wang et al., 2025). Among human factors influencing this ability, driving experience has emerged as the most extensively studied variable (Sagberg & Bjørnskau, 2006; Scialfa et al., 2012; Crundall et al., 2012; Bonfiglio et al., 2014; Vlakveld, 2014; Crundall, 2016), with the role of (self-reported) confidence and its miscalibration (relative to performance outcomes) seldom being investigated.

Literature regarding the consistency between subjectively and objectively measured hazard perception abilities has yielded mixed findings (Farrand and McKenna, 2001; Martinussen et al., 2017; Åbele et al., 2018). For instance, Farrand and McKenna (2001) found no association between self-reported questionnaires and behavioral performance, suggesting an independence between questionnaire and response latency measure. In contrast, Åbele and colleagues (2018) reported that young drivers who responded in time to visible hazards also scored higher on both subjective self-assessments and objective measures of HPS. Recent research has further investigated whether confidence-related bias may directly influence hazard perception performance (Sun et al., 2024; Hu et al., 2025). Sun and colleagues (2024), using a static-image paradigm of the hazard perception test, showed that overconfident drivers exhibit slower response times compared to controls. Nonetheless, static images provide a less ecologically valid representation of real-world driving. Addressing the limitations of this previous research, Hu, Sun, and Cheng (2025) used dynamic video stimuli and

classified drivers into several confidence groups (rather than using a binary classification adopted by Sun and colleagues, 2024), based on their self-rated scores on the Hazard Perception Questionnaire (HPQ; White et al. 2011). Their results showed that underconfident drivers reported longer response times than both moderately and very confident drivers. It follows that both over- and under- confidence can impair hazard perception, highlighting the influence of confidence bias. Since previous studies primarily define confidence through self-reported questionnaires, the direct comparison between the ability to judge the correctness of one's task performance and the actual performance outcomes remain overlooked. Addressing this gap, this study proposes a performance-based approach investigating the role of confidence bias, defined as a continuous variable, in shaping both hazard perception accuracy and subjective risk evaluation of road scenarios.

3.1.2 Individual differences in Environmental Sensitivity

Individuals' responses may be shaped by their sensitivity to surrounding factors, with some individuals being more or less reactive to the same environmental conditions (Belsky & Pluess, 2009; Aron et al., 2012; Pluess, 2015). This manuscript took into consideration the role of Environmental Sensitivity, an individual trait that embodies genetic, neurophysiological, and behavioral differences in how people perceive and respond to stimuli (Pluess, 2015). Its theoretical foundation is grounded in three conceptual frameworks, comprising Aron and Aron's sensory processing sensitivity (1997), Belsky's differential susceptibility model (1997, 2009), and the biological sensitivity to context theory (Boyce & Ellis, 2005; Ellis et al., 2011). Recently, it has been conceptualized as an adaptive trait that allows individuals to respond to both negative and positive environmental conditions, serving as the foundation for human adaptability and neuroplasticity (Pluess et al., 2018). This evolution is reflected in the measurement of this trait, from the original Highly Sensitive Person scale (HSP, Aron & Aron, 1997) to more recent adaptation (Pluess et al., 2023). Although the original scale covers broad domains, it has been linked to heightened perceptual awareness and responsiveness to threats. As evidenced by findings from Rubaltelli and colleagues (2018), higher HSP scores predicted greater risk perception and a stronger psychophysiological response to threatening visual stimuli. To the best of our

knowledge, its relationship with subjective risk estimation of potential road hazard has yet to be explored.

3.1.3 Overconfidence and Driving Experience

Overconfidence and driving experience could demonstrate an association, though directionality remains exploratory due to conflicting evidence in prior research. On the one hand, according to the zero-risk theory (Näätänen & Summala, 1974), as drivers develop greater proficiency, the more they adapt to risks on the road. This exaggerated sense of control and overconfidence in one's abilities represent a significant risk factor for traffic safety, as it reduces the driver's risk perception (Summala, 1988). Other studies have further corroborated that driver's experience is positively correlated with increased confidence in their driving skills and negatively associated with safety concerns (Lajunen & Summala, 1995). On the other hand, Deery (1999) theorizes his model by emphasizing how young novice drivers often overestimate their driving abilities while underestimating the risks associated with hazardous situations. This overconfidence may be a key factor contributing to their higher involvement in accidents compared to experienced drivers.

3.1.4 The current study

By considering confidence bias as a continuous predictor, this study attempts to examine its impact on hazard perception accuracy and subjective risk estimation, accounting for environmental factors (i.e. lighting conditions and hazard presence). It also takes into account individual differences, such as Environmental Sensitivity and driving experience. Based on the background discussed above, the following research aims are addressed.

- (1) We examine the role of confidence bias, lighting conditions (daytime, nighttime), and hazard presence (hazard, safe) on hazard perception performance, specifically on response accuracy (RQ1a) and reaction times (RQ1b).
- (2) We further evaluate how these same factors influence subjective risk estimation for potential traffic hazards (RQ2a), and explore if Environmental Sensitivity

(HSP-12, Pluess et al., 2023) relates with individual differences in rating risk (RQ2b).

(3) Finally, we examine the association between overconfidence and driving experience (RQ3).

3.2 Method

3.2.1 Participants

Sixty-five participants took part in the experiment (32 female, 1 non-binary and 32 male; age range 20-35 years, $M = 26.31$ years, $SD = 4.33$ years). They obtained a car driver's license from 1-17 years ($M = 7.32$ years; $SD = 4.21$ years) and were student volunteers from the D'Annunzio University of Chieti-Pescara.

3.2.2 Materials

Self-Report Questionnaires

Socio-demographic information (including age, gender, level of education) and driving experience were collected via questionnaire. Driving experience refers to self-reported measures of road exposure, including years since obtaining a license, driving frequency, and annual kilometers driven (see Castro et al., 2020).

Self-reported hazard perception skill was assessed using the Hazard Perception Questionnaire (HPQ-6) (White et al., 2011). Participants were asked to compare their ability in hazard perception to those of an average driver (e.g., item 1: "Compared to an average driver, how skilled are you at spotting hazards quickly?"). A 7-point Likert scale was used, ranging from 1 ("much less") to 7 ("much more"), with a midpoint of 4 ("the same"). In this study, the internal consistency reliability of the Hazard Perception Questionnaire was 0.88.

The 12-item Highly Sensitive Person scale (HSP, Pluess et al., 2023) was used to measure the participants' Environmental Sensitivity. They were asked to indicate their level of agreement with 12 items (e.g., Item 2: "Are you easily overwhelmed by things like bright lights, strong smells, coarse fabrics, or sirens close by?") on a 7-point Likert

scale ranging from 1 ("Not at all") to 7 ("Extremely"). The internal consistency reliability of the Highly Sensitive Person Scale was 0.83.

Video-based Hazard Perception Task (HPT)

The video-based Hazard Perception Task (Wetton et al., 2011) was programmed in E-Prime 3 (Psychology Software Tools, 2017; <https://pstnet.com/products/e-prime/>), and contained 80 video clips selected from the *Road Hazard Stimuli* dataset (for details see Song et al., 2024). Each video showed a road scenario filmed using front-facing dashcams. The inclusion criteria for the videos required: an average duration of approximately 8 seconds ($M = 7.78$ s; $SD = 1.25$), adherence to right-hand traffic conventions (in line with Italian driving regulations), visual stability of the dashcam, absence of distracting text overlays, and a resolution of 1280×720 pixels (30 frames per second). Videos were selected, such that half depicted a hazardous event and half did not, balanced for daytime and nighttime conditions. Hazards include pedestrians, vehicles, animals, and obstacles in a range of environments (e.g., city streets, highways), weather (e.g., sunny, rainy, or snowy), and lighting conditions (daytime or nighttime). The clips were presented in random order. The videos have been used and validated in previous studies (Song & Wolfe, 2024; Guidi et al., 2024). Detailed categorical and temporal information on the Hazard Perception Task stimuli is provided in Supplementary Materials (Table A2).

Before starting the task, the following definition of hazard was given, as “any object, situation, occurrence or combination of these that introduce the possibility of the individual road user experiencing harm should be included [...]. Harm may include damage to one’s vehicle, injury to oneself, damage to another’s property, or injury to another person” (Haworth et al. 2000, p. 3; Borowsky et al., 2010). During the task, participants observed driving scenes recorded from a driver’s perspective, imagining themselves as actors in the clips. They were instructed to press the spacebar only when they identified a hazardous situation that required action to avoid a crash (response), and to withhold responding otherwise. The video would stop in case of responding, and therefore were requested to respond only once to the hazard they identified.

Response outcomes (whether participants pressed the spacebar or not) and Response Times to hazard (defined as the moment at which the spacebar was pressed to respond) were recorded for each trial of the Hazard Perception Task.

After each video, participants provided trial-by-trial confidence judgments and subjective risk estimation (for details, see the following two sections). An example of a trial from this novel adaptation of the Hazard Perception Task is shown in Figure 3.1.

Driver's Confidence Judgment Test

In this study, the driver's confidence judgment measure consisted of one question after each video trial of the Hazard Perception Task (HPT). Specifically, participants self-reported the confidence in their perceptual performance accuracy, by answering the question: "*How confident are you that your decision [to press or not the spacebar] is correct? Please give a rate on a scale from 1 to 7*".

Subjective Risk Estimation Test

Subjective risk estimation, designed to evaluate the perceived level of danger in each road scenario, was assessed immediately after each video trial of the Hazard Perception Task (HPT). Participants were prompted to answer the question: "*How dangerous does this situation seem to you?*" by selecting a response on a 7-point Likert scale ranging from 1 ("Not hazardous at all") to 7 ("Very hazardous") (see Castro et al., 2020). This measure is consistent with the definition of risk perception, as "a subjective judgment that categorizes an upcoming event as potentially dangerous" (Knuth et al., 2014).

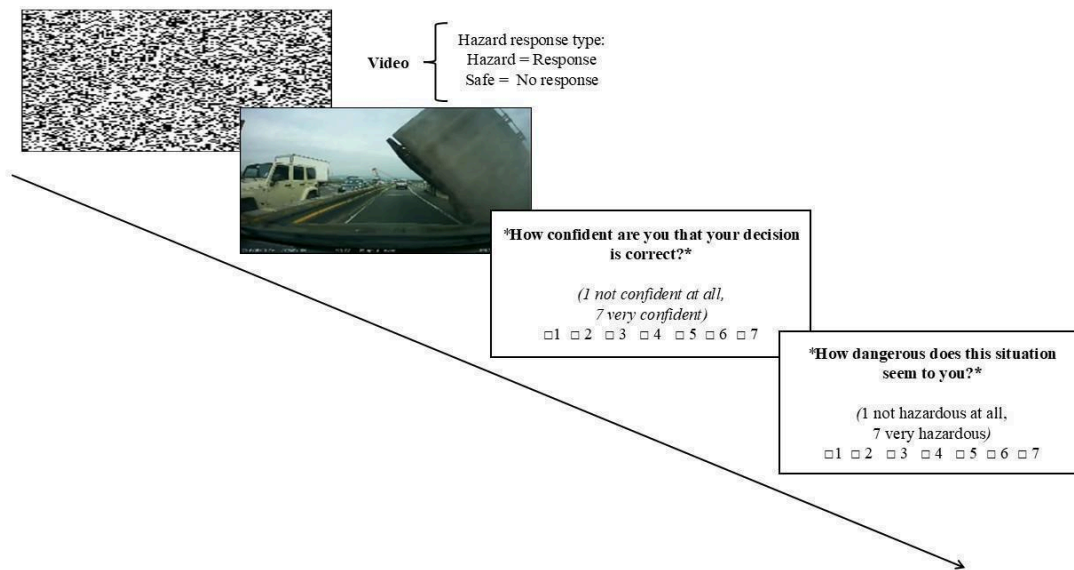


Figure 3.1 Example of a single trial of the task. The timeline shows: the initial mask slide (duration 350 ms), followed by a road traffic video clip (approximately 8 seconds) during which participants could respond or not; after each video, they provided trial-by-trial rating of confidence judgment and subjective risk estimation on a scale from 1 to 7.

Balloon Analogue Risk Task (BART)

BART was used as a control variable to eliminate the effects of risk-taking behavior (Lejuez et al., 2002). Participants inflated a virtual computer-based balloon for earning points, with a random bursting threshold. They could stop inflating at any time to secure accumulated points. The task included one practice trial and ten critical trials. Risk tolerance was calculated as the percentage of burst balloons, which suggests risk-taking behavior. Higher percentages reflected a greater willingness to take risks, offering insights into individual differences in decision-making under uncertainty.

3.2.3 Experimental Design

A 2 (hazard presence: hazard, safe) x 2 (lighting conditions: daytime, nighttime) within subject factorial design was used, including confidence bias as a continuous between subject variable. The dependent variables were: (1) Response accuracy (binary

outcome) in the HPT, (2) Reaction Times to hazard during the HPT, (3) Subjective risk estimation (Likert scale).

3.2.4 Procedure

Upon arriving at the laboratory, participants received information about the experiment and signed an informed consent form. They completed three questionnaires: (1) demographic and driving experience; (2) HPQ-6; and (3) HSP-12. Then, participants were instructed to sit 80 cm from the screen of a 15.6" HP Windows laptop, where the experimenter provided instructions for the behavioral tasks. E-Prime 3 software (<https://pstnet.com/products/e-prime/>) was used to present the written instructions and record participants' behavioral responses. First, participants performed the Balloon Analog Risk Task (BART). Next, the instructions about how to proceed during the video-based Hazard Perception Task (HPT) and how to rate each trial in terms of confidence judgment and subjective risk estimation were given. Questionnaires were administered before the Video-based Hazard Perception Task to limit possible response bias created by driving scenarios (see Abele et al., 2018). The entire experiment lasted approximately 35 minutes.

3.2.5 Data Analysis

Hazard Perception Task (HPT)

A total of 27 trials were removed due to a wrong keyboard case pressed during the task. Response accuracy was scored binary, marking the response as “correct” in the case of hazard identification (in the hazard videos) and omission of response (in the safe videos); misses and false alarms were marked as “incorrect”. For the hazard videos, Reaction Times (RTs) were calculated as the difference between the participant’s response time and the hazard onset time. The hazard onset time, as defined and pre-annotated in the *Road Hazard Stimuli* dataset (Song et al., 2024), refers to the first visible deviation point from the normal state at which it could be detected. It varied randomly across videos. A positive difference between the participant's response time and hazard onset indicates a delayed response, a negative difference reflects anticipation of the hazard, and finally, a value of zero indicates a response coinciding with the hazard appearance.

Generalized linear mixed-effects models (GLMMs) and Linear Mixed Models (LMMs) were used to investigate the effects of hazard presence (hazard vs. safe), lighting condition (daytime vs. nighttime), and confidence bias on the dependent variables (response accuracy, RTs, subjective risk estimation). Response accuracy was examined using a GLMM with a binomial distribution and logit link function, whereas RTs (on hazard trials only) and subjective risk estimation were analyzed using LMMs. Random intercepts were included to account for between-participant variability. Confidence bias was treated as a continuous predictor. For response accuracy (binomial outcome), model based predicted probabilities were estimated on the response scale, and findings are reported as percentage-point (*pp*) differences relative to a baseline condition defined as safe daytime trials with confidence bias fixed at its sample mean. For RTs and subjective risk estimation, effects are reported as differences in estimated marginal means.

Confidence Bias

Confidence bias was determined as the deviation between participants' (self-reported) confidence judgments regarding the correctness of one's task performance and their (objective) performance in the Hazard Perception Task. Adopting the method proposed by Sun and colleagues (2024), this measure was calculated, for each participant, as the difference between the percentage of confidence judgments and the percentage of response accuracy. A difference of zero indicates perfect calibration (i.e., the ability to correctly judge one's task performance), while positive values reflect overconfidence. For descriptive analyses, the sample ($N = 65$) was classified into two groups: twenty-nine overconfident individuals (44.6%; positive deviation), and thirty-six not overconfident (55.4%; zero or negative deviation). In the main analysis, confidence bias was included as a continuous predictor. Formulas used to compute hazard perception accuracy, confidence judgments, confidence bias, and group classification are provided in the Supplementary Materials (A5).

Driving Experience

Based on the criteria established by Castro and colleagues (2020), the sample was classified according to the driving experience. Twenty-six were considered experienced

drivers (40%), as they had held a car driver's license for more than three years, drove at least twice a week, and covered more than 10,000 km per year. The remaining thirty-nine participants (60%), who did not meet these criteria, were categorized as novices. A Chi-square test of independence was performed to explore the association between confidence bias-based groups (overconfident vs. not overconfident) and driving experience-based groups (novice vs. experienced drivers).

Availability of data and materials

The datasets and materials used and/or analyzed during the current study are available on the Open Science Framework (OSF) at https://osf.io/jb72y/?view_only=2728ddd1a9ce457b87790de7026e99ee, ensuring adherence to data transparency practices.

Code availability

The code utilized for the analysis in this study is available on the Open Science Framework (OSF) at https://osf.io/jb72y/?view_only=2728ddd1a9ce457b87790de7026e99ee, ensuring adherence to data transparency practices.

3.3 Results

3.3.1 Descriptive statistics

Table 3.1 shows the descriptive analysis of the total sample (N = 65) divided in two confidence bias-based groups. It includes demographic data (age and driving years), mean scores on the HPQ-6 questionnaires and HPS-12 scale, mean percentage of self-reported confidence judgments and mean percentage of exploded balloons in the BART.

	Overconfident Group		Not Overconfident Group	
	N = 29		N = 36	
	Mean (SD)	%	Mean (SD)	%
Age	26.29 (4.37)		26.32 (4.35)	
Driving Years	7.45 (4.14)		7.21 (4.34)	
HPQ-6	5.28 (0.83)		4.96 (0.87)	
HSP-12	4.43 (1.10)		4.66 (0.79)	
Confidence Judgments		95.26		86.57
BART (burst balloons)		23.87		27.94

Table 3.1 Comparison of descriptive statistics between groups.

3.3.2 Behavioral data analysis

Table 3.2 summarizes the results of the (G)LMM analyses. The main text only discusses the main significant effects. Full fixed-effects estimates are provided in the Supplementary Materials (Table B1, B2, B3).

Response Accuracy: Confidence bias showed a significant effect on accuracy ($\beta = -0.04$, $SE = 0.01$, $t = -4.48$, $p < .001$), with positive values of confidence bias (i.e., participants whose subjective confidence judgments exceeded their actual performance, deviation > 0) associated with a lower probability of a correct response. Also, as illustrated in Figure 3.2A, a significant hazard presence x lighting conditions interaction emerged ($p < .001$). The baseline performance, defined as predicted accuracy in safe daytime trials with confidence bias fixed at its mean value, was estimated at 96.4%. During daytime, accuracy was slightly lower in hazardous scenarios compared to safe ones (-1.8 percentage points, $p = .024$). In contrast, during nighttime conditions hazard trials were associated with higher accuracy than safe trials (+8.9 percentage points, $p < .001$). In safe trials, accuracy decreased at night relative to daytime (-9.7 percentage points, $p < .001$), but no significant day-night difference was observed in hazardous

trials ($p = .21$). No interactions involving confidence bias reached statistical significance.

Reaction Times (RTs): A significant main effect of lighting conditions was found ($\beta = -98.82$, $SE = 17.41$, $t = -5.67$, $p < .001$). Estimated marginal means confirmed slower responses during nighttime scenarios ($M = 934.8$ ms, 95% CI [841.6, 1028.0]), compared to daytime ($M = 737.1$ ms, 95% CI [643.8, 830.5]). Confidence bias did not show a significant effect on RTs ($p = .45$).

Subjective Risk Estimation: A significant hazard presence $\times$ lighting conditions interaction emerged ($p < .001$; Figure 3.2B), suggesting that the effect of hazard presence on risk ratings differed between daytime and nighttime. In the baseline condition (safe daytime trials), the estimated marginal mean risk score was 1.74 (Likert) points (95% CI [1.58, 1.89]). Relative to this baseline, subjective risk estimation was higher in hazardous scenarios presented during daytime ($+3.79$, $p < .001$). Safe scenarios during nighttime showed a smaller but significant increase in rating risk ($+0.63$, $p < .001$). Hazardous nighttime scenarios were associated with the largest increase in rating risk relative to the baseline ($+3.82$, $p < .001$). Confidence bias did not exhibit a significant main effect on subjective risk estimation ($p = .68$). Although the interaction between confidence bias and hazard presence reached statistical significance ($p < .001$), conditional slopes analysis showed no reliable association between confidence bias and risk estimation in either hazardous trials (slope = 0.009, 95% CI [-0.010, 0.029]) or safe ones (slope = -0.0017, 95% CI [-0.037, 0.002]). No interaction between confidence bias and lighting condition emerged ($p = .86$).

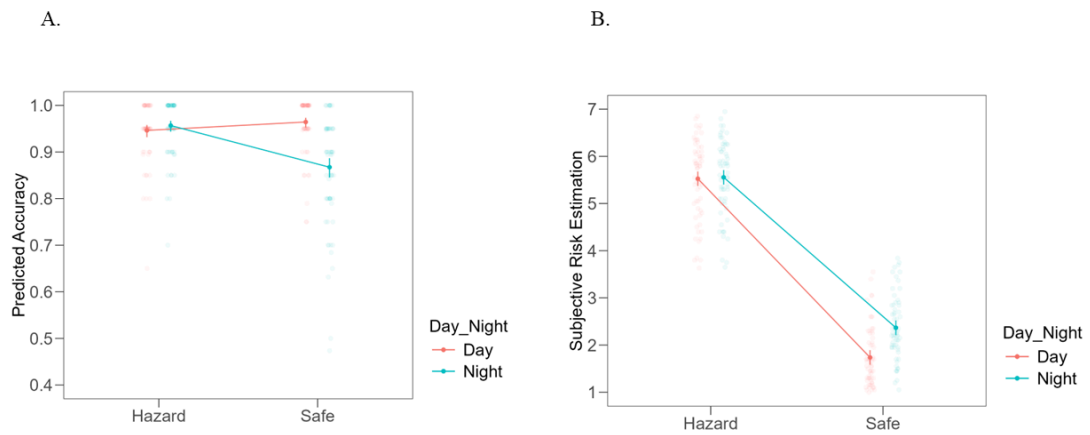


Figure 3.2. *Model-based estimates from mixed effects model as a function of hazard presence and lighting conditions (95% CIs). Panels show: (A) predicted accuracy from the generalized linear mixed-effects model (GLMM), and (B) estimated marginal means of risk ratings from the linear mixed-effects model (LMM).*

	Accuracy (pp)	RTs (ms)	Risk Estimation (Likert point)
Confidence bias	↓ ***	ns	ns
Nighttime (Safe trials)	- 9.7***	—	+0.63***
Nighttime (Hazard trials)	—	+197.7***	—
Hazard (Daytime)	- 1.8*	—	+3.79***
Hazard (Nighttime)	+8.9***	—	+3.82***

Table 3.2 *Model based effects relative to the safe-daytime baseline (except for confidence bias, which reflects the main effect of the continuous predictor). Findings are reported as percentage point (pp) difference for accuracy, milliseconds for RTs, and (Likert) point differences for subjective risk estimation. Significance: *: $p < .05$; **: $p < .01$; ***: $p < .001$. Dashes (—) indicate effects not estimated. Downward arrow (↓) indicates a negative effect.*

Environmental Sensitivity and Subjective Risk Estimation: Pearson's correlation analysis revealed a significant positive correlation ($r = 0.028$, $p = 0.047$, 95% CI [0.000, 0.055]) between the score on Highly Sensitive Person scale (HSP; Pluess et al., 2023) and risk ratings in driving scenarios.

Overconfidence and Driving Experience: Chi-Squared test of independence revealed a statistically non-significant association between these variables, $\chi^2(1, 65) = 0.508$, $p = 0.47$, indicating that in our sample the distribution of overconfidence does not significantly differ based on driving expertise.

3.4 Discussion

The present study explored how confidence bias and environmental factors influence perception in driving, with regard to both subjective experience of risk and performance in a video-based Hazard Perception Task (HPT). First, a key contribution of this work is conceptualizing confidence bias as the difference between self-reported ratings of their task accuracy versus their objective task accuracy on each trial of HPT. Unlike previous studies, which relied primarily on self-report questionnaires representing an off-line global self-monitoring of driving abilities (Farrand & McKenna, 2001; Martinussen et al., 2017; Åbele et al., 2018; Händel et al., 2020; Hu et al., 2025), this work captures on-line trial-by-trial confidence judgments on correctness of one's own performance (Saraç & Karakelle, 2012). Second, by treating confidence bias as a continuous variable, it advances previous group-based studies (Sun et al., 2024; Hu et al., 2025), allowing us to gather individual variability.

Task-level analysis of hazard perception performance showed that confidence bias was a significant predictor of response accuracy. Overconfident participants were associated with a lower probability of correct response in identifying hazards. This pattern supports the theoretical view that judgment miscalibration (Moore & Healy, 2008; Horrey et al., 2015) may undermine hazard identification in dynamic traffic environments. The effect of confidence bias on reaction times to road hazards was not observed, differing from previous studies based on confidence groups (Sun et al., 2024; Hu et al., 2025). This discrepancy likely reflects methodological differences in how overconfidence was operationalised across studies. Future work may benefit directly comparing questionnaire- and performance-based measures of confidence, along with continuous- and group-based approaches, within the same experimental paradigm. Beyond overconfidence, environmental factors significantly shaped hazard perception performance. As expected based on previous studies (Garay et al., 2004; Konstantopoulos et al., 2010; Asadamraji et al., 2019; Wang et al., 2025), performance was affected by both hazard presence and lighting conditions (see Figure 3.2A). Nighttime scenarios significantly elicited longer reaction times to hazards compared to daytime, which confirms that low luminance can impair visual perception (Wang et al.,

2025). Accuracy in hazardous trials remained relatively high across lighting conditions, while it declined during nighttime in safe trials.

Subjective risk estimation was primarily determined by the presence of hazard, and significantly modulated by lighting conditions (Deery, 1999; Horswill & McKenna, 2004; see Figure 3.2B). Confidence bias did not affect estimation of the perceived level of risk in a driving scenario. It's important to point out that unlike previous evidence linked overconfidence to underestimation of personal crash risk involvement (Matthews & Moran, 1986; Gregersen, 1996; Deery, 1999; Sümer et al., 2006; McKenna & Horswill, 2006), the present study assessed context-specific risk ratings rather than self-referential risk. Within this framework, our findings would suggest that risk ratings of traffic scenarios depend more on context than on individual characteristics. Future research should further study how people evaluate the riskiness of scenes along with their own personal risk, and whether overconfidence affects them.

Environmental Sensitivity (HSP scale) demonstrated a weak but significant correlation with subjective risk estimation of road scenarios. Specifically, individuals who scored higher on the Highly Sensitive Person scale (Pluess et al., 2023), tended to report slightly greater perceived risk. This finding is consistent with the literature suggesting that heightened environmental sensitivity reflects increased responsiveness to external cues (Pluess, 2015) and threat awareness (Rubaltelli et al., 2018). Despite the small effect size, this finding warrants further investigations.

In this task, no significant relationship between overconfidence and driving experience was found. Notably, our sample did not include professional drivers, limiting generalizability to more advanced proficiency. Theoretical perspectives remain divided, with some suggesting greater experience increases overconfidence and reduces risk perception (Näätänen & Summala, 1974; Lajunen & Summala, 1995), while others highlight overconfidence in novice drivers (Deery, 1999). The lack of significance highlights the need for further research, including studies on professional drivers, to better understand how experience influences overconfidence.

3.4.1 Limitations and Implications

The present research has some limitations. First, the absence of differentiation between latent and immediate hazards in the video stimuli used. Our definition of hazards did not account for latent hazards, which could have limited the generalizability of findings to scenarios involving less obvious dangers (Song et al., 2024). Second, hazards occurred in 50% of the stimulus set, which does not reflect the probabilities of occurrence in the real world. This manipulation may influence participants' expectations and responses. Future studies should address these limitations to improve the design of stimuli in the HPT.

This study has practical applications. Findings indicate that confidence bias can impair response accuracy in perceiving hazards in driving scenarios, highlighting the key relevance of confidence calibration in potentially real-world driving performance. Although Hazard Perception Task is already implemented in driver licensing systems across several countries as an early indicator of driving skill (e.g., Crundall et al., 2021; Wetton et al., 2011), confidence measurement in judging task performance should be integrated into driving tests. Based on studies showing that feedback-based training can effectively reduce bias and overconfidence in self-assessments of driving abilities (e.g., Horswill et al., 2017), training programs targeting overconfidence could contribute to safe driving.

3.5 Conclusion

This study reveals insight regarding drivers' hazard perception performance and subjective risk estimation based on confidence bias and environmental conditions in a video-based hazard perception task. Overconfidence (i.e., positive values of confidence bias) was associated with a reduced response accuracy in perceiving hazards, but it did not affect reaction times. Also, hazard perception performance and subjective risk estimation resulted from the interaction between the presence of hazard and lighting conditions. Confidence bias did not influence subjective risk estimation, suggesting that risk ratings of traffic scenarios may rely more on environmental cues. A weak positive association between Environmental Sensitivity (HSP scale) and subjective risk estimation suggests that this trait may amplify responsiveness to potential threats in

traffic context. An absent link emerged between overconfidence and driving experience. The results suggest that integrating the measurement of confidence in performance accuracy into driver training and testing could improve awareness and safety.

References

- Åbele, L., Haustein, S., Møller, M., & Martinussen, L. M. (2018). Consistency between subjectively and objectively measured hazard perception skills among young male drivers. *Accident Analysis & Prevention*, 118, 214–220. <https://doi.org/10.1016/j.aap.2018.02.022>
- Aron, E. N., & Aron, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. *Journal of Personality and Social Psychology*, 73(2), 345–368. <https://doi.org/10.1037/0022-3514.73.2.345>
- Aron, E. N., Aron, A., & Jagiellowicz, J. (2012). Sensory Processing Sensitivity: A Review in the Light of the Evolution of Biological Responsivity. *Personality and Social Psychology Review*, 16(3), 262–282. <https://doi.org/10.1177/1088868311434213>
- Asadamraji, M., Saffarzadeh, M., Ross, V., Borujerdian, A., Ferdosi, T., & Sheikholeslami, S. (2019). A novel driver hazard perception sensitivity model based on drivers' characteristics: A simulator study. *Traffic Injury Prevention*, 20(5), 492–497. <https://doi.org/10.1080/15389588.2019.1607971>
- Baker, M., & Wurgler, J. (2013). Chapter 5 - Behavioral Corporate Finance: An Updated Survey. In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.), *Handbook of the Economics of Finance* (Vol. 2, pp. 357–424). Elsevier. <https://doi.org/10.1016/B978-0-44-453594-8.00005-7>
- Belsky, J. (1997). Variation in susceptibility to environmental influence: An evolutionary argument. *Psychological Inquiry*, 8(3), 182–186. https://doi.org/10.1207/s15327965pli0803_3
- Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin*, 135(6), 885–908.

<https://doi.org/10.1037/a0017376>

- Bonfiglio, D. J., Lee, Y.-C., LaVoie, N., & Parker, J. (2014). Progression of hazard perception knowledge, licensure status and driving experience among teen drivers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 2136–2140. <https://doi.org/10.1177/1541931214581449>
- Borowsky, A., Oron-Gilad, T., & Parmet, Y. (2010). The role of driving experience in hazard perception and categorization: A traffic-scene paradigm. *World Academy of Science, Engineering and Technology*, 66, 305–309. <https://www.scopus.com/pages/publications/79951530926>
- Boyce, W. T., & Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary-developmental theory of the origins and functions of stress reactivity. *Development and Psychopathology*, 17(2), 271–301. <https://doi.org/10.1017/S0954579405050145>
- Brus, J., Aebersold, H., Grueschow, M., & Polania, R. (2021). Sources of confidence in value-based choice. *Nature Communications*, 12(1), 7337. <https://doi.org/10.1038/s41467-021-27618-5>
- Cao, S., Samuel, S., Murzello, Y., Ding, W., Zhang, X., & Niu, J. (2022). Hazard Perception in Driving: A Systematic Literature Review. *Transportation Research Record*, 2676(12), 666–690. <https://doi.org/10.1177/03611981221096666>
- Castro, C., Muela, I., Doncel, P., & García-Fernández, P. (2020). Hazard Perception and Prediction test for walking, riding a bike and driving a car: “Understanding of the global traffic situation.” *PLoS ONE*, 15(10). <https://doi.org/10.1371/journal.pone.0238605>
- Cheng, A. S. K., Ng, T. C. K., & Lee, H. C. (2011). A comparison of the hazard

- perception ability of accident-involved and accident-free motorcycle riders. *Accident Analysis & Prevention*, 43(4), 1464–1471. <https://doi.org/10.1016/j.aap.2011.02.024>
- Chua, E. F., & Solinger, L. A. (2015). Building metamemorial knowledge over time: Insights from eye tracking about the bases of feeling-of-knowing and confidence judgments. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01206>
- Chuang, W.-I., & Lee, B.-S. (2006). An empirical evaluation of the overconfidence hypothesis. *Journal of Banking & Finance*, 30(9), 2489–2515. <https://doi.org/10.1016/j.jbankfin.2005.08.007>
- Croskerry, P., & Norman, G. (2008). Overconfidence in Clinical Decision Making. *The American Journal of Medicine*, 121(5), S24–S29. <https://doi.org/10.1016/j.amjmed.2008.02.001>
- Crundall, D. (2016). Hazard prediction discriminates between novice and experienced drivers. *Accident Analysis & Prevention*, 86, 47–58. <https://doi.org/10.1016/j.aap.2015.10.006>
- Crundall, D., Chapman, P., Phelps, N., & Underwood, G. (2003). Eye movements and hazard perception in police pursuit and emergency response driving. *Journal of Experimental Psychology: Applied*, 9(3), 163–174. <https://doi.org/10.1037/1076-898X.9.3.163>
- Crundall, D., Chapman, P., Trawley, S., Collins, L., van Loon, E., Andrews, B., & Underwood, G. (2012). Some hazards are more attractive than others: Drivers of varying experience respond differently to different types of hazard. *Accident Analysis & Prevention*, 45, 600–609. <https://doi.org/10.1016/j.aap.2011.09.049>
- Crundall, D., van Loon, E., Baguley, T., & Kroll, V. (2021). A novel driving

- assessment combining hazard perception, hazard prediction and theory questions. *Accident Analysis & Prevention*, 149, 105847. <https://doi.org/10.1016/j.aap.2020.105847>
- Deery, H. A. (1999). Hazard and Risk Perception among Young Novice Drivers. *Journal of Safety Research*, 30(4), 225–236. [https://doi.org/10.1016/S0022-4375\(99\)00018-3](https://doi.org/10.1016/S0022-4375(99)00018-3)
- Efklides, A., & Misailidi, P. (2010a). Introduction: The Present and the Future in Metacognition. In A. Efklides & P. Misailidi (Eds.), *Trends and Prospects in Metacognition Research* (pp. 1–18). Springer US. https://doi.org/10.1007/978-1-4419-6546-2_1
- Efklides, A., & Misailidi, P. (Eds.). (2010b). *Trends and Prospects in Metacognition Research*. Springer US. <https://doi.org/10.1007/978-1-4419-6546-2>
- Ellis, B. J., Boyce, W. T., Belsky, J., Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2011). Differential susceptibility to the environment: An evolutionary–neurodevelopmental theory. *Development and Psychopathology*, 23(1), 7–28. <https://doi.org/10.1017/S0954579410000611>
- E-Prime® Stimulus Presentation Software | Psychology Software Tools*. (2017, September 11). Psychology Software Tools | Solutions for Research, Assessment, and Education. <https://pstnet.com/products/e-prime/>
- Evans, T., Stuckey, R., & Macdonald, W. (2020). Young drivers' perceptions of risk and difficulty: Day versus night. *Accident Analysis & Prevention*, 147, 105753. <https://doi.org/10.1016/j.aap.2020.105753>
- Farrand, P., & McKenna, F. P. (2001). Risk perception in novice drivers: The relationship between questionnaire measures and response latency.

- Transportation Research Part F: Traffic Psychology and Behaviour*, 4(3), 201–212. [https://doi.org/10.1016/S1369-8478\(01\)00024-9](https://doi.org/10.1016/S1369-8478(01)00024-9)
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fetsch, C. R., Kiani, R., & Shadlen, M. N. (2014). Predicting the Accuracy of a Decision: A Neural Mechanism of Confidence. *Cold Spring Harbor Symposia on Quantitative Biology*, 79, 185–197. <https://doi.org/10.1101/sqb.2014.79.024893>
- Fisher, D. L., Pollatsek, A. P., & Pradhan, A. (2006). Can novice drivers be trained to scan for information that will reduce their likelihood of a crash? *Injury Prevention: Journal of the International Society for Child and Adolescent Injury Prevention*, 12 Suppl 1(Suppl 1), i25–29. <https://doi.org/10.1136/ip.2006.012021>
- Fleming, S. M. (2024). Metacognition and Confidence: A Review and Synthesis. *Annual Review of Psychology*, 75(Volume 75, 2024), 241–268. <https://doi.org/10.1146/annurev-psych-022423-032425>
- Garay, L., Fisher, D. L., & Hancock, K. L. (2004). Effects of Driving Experience and Lighting Condition on Driving Performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48(19), 2290–2294. <https://doi.org/10.1177/154193120404801918>
- Gregersen, N. P. (1996). Young drivers' overestimation of their own skill—An experiment on the relation between training strategy and skill. *Accident Analysis & Prevention*, 28(2), 243–250. [https://doi.org/10.1016/0001-4575\(95\)00066-6](https://doi.org/10.1016/0001-4575(95)00066-6)

- Groeger, J. A. (2013). *Understanding Driving: Applying Cognitive Psychology to a Complex Everyday Task*. Routledge. <https://doi.org/10.4324/9780203769942>
- Guidi, S., Kosovicheva, A., & Wolfe, B. (2024). Seeing the truck, but missing the cyclist: Effects of blur on duration thresholds for road hazard detection. *Cognitive Research: Principles and Implications*, 9(1), 32. <https://doi.org/10.1186/s41235-024-00557-7>
- Händel, M., de Bruin, A. B. H., & Dresel, M. (2020). Individual differences in local and global metacognitive judgments. *Metacognition and Learning*, 15(1), 51–75. <https://doi.org/10.1007/s11409-020-09220-0>
- Haworth, N., Symmons, M., & Kowadlo, N. (2000). *Hazard perception by inexperienced motorcyclists*. <https://trid.trb.org/View/732859>
- Hills, B. L. (1980). Vision, Visibility, and Perception in Driving. *Perception*, 9(2), 183–216. <https://doi.org/10.1068/p090183>
- Horrey, W. J., Lesch, M. F., Mitsopoulos-Rubens, E., & Lee, J. D. (2015). Calibration of skill and judgment in driving: Development of a conceptual framework and the implications for road safety. *Accident Analysis & Prevention*, 76, 25–33. <https://doi.org/10.1016/j.aap.2014.12.017>
- Horswill, M. S., Garth, M., Hill, A., & Watson, M. O. (2017). The effect of performance feedback on drivers' hazard perception ability and self-ratings. *Accident Analysis & Prevention*, 101, 135–142. <https://doi.org/10.1016/j.aap.2017.02.009>
- Horswill, M. S., Hill, A., & Wetton, M. (2015). Can a video-based hazard perception test used for driver licensing predict crash involvement? *Accident Analysis & Prevention*, 82, 213–219. <https://doi.org/10.1016/j.aap.2015.05.019>

- Horswill, M. S., & McKenna, F. P. (1999). The Effect of Perceived Control on Risk Taking. *Journal of Applied Social Psychology*, 29(2), 377–391. <https://doi.org/10.1111/j.1559-1816.1999.tb01392.x>
- Horswill, M. S., & McKenna, F. P. (2004). Drivers' hazard perception ability: Situation awareness on the road. In S. Banbury & T. S. (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 155–175). Ashgate. <https://centaur.reading.ac.uk/14289/>
- Hu, W., Sun, L., & Cheng, L. (2025). Effects of confidence level and hazard type on the visual search patterns and hazard response times of young drivers. *Traffic Injury Prevention*, 1–7. <https://doi.org/10.1080/15389588.2025.2497521>
- Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1322–1337. <https://doi.org/10.1098/rstb.2012.0037>
- Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, 84(6), 1329–1342. <https://doi.org/10.1016/j.neuron.2014.12.015>
- Knuth, D., Kehl, D., Hulse, L., & Schmidt, S. (2014). Risk perception, experience, and objective risk: A cross-national study with European emergency survivors. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, 34(7), 1286–1298. <https://doi.org/10.1111/risa.12157>
- Konstantopoulos, P., Chapman, P., & Crundall, D. (2010). Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore drivers' eye movements in day, night and rain driving. *Accident*

- Analysis & Prevention*, 42(3), 827–834.
<https://doi.org/10.1016/j.aap.2009.09.022>
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 107–118. <https://doi.org/10.1037/0278-7393.6.2.107>
- Kumar, J., & Prince, N. (2023). Overconfidence Bias in Investment Decisions: A Systematic Mapping of Literature and Future Research Topics. *FIIB Business Review*, 23197145231174344. <https://doi.org/10.1177/23197145231174344>
- Lajunen, T., & Summala, H. (1995). Driving experience, personality, and skill and safety-motive dimensions in drivers' self-assessments. *Personality and Individual Differences*, 19(3), 307–318.
[https://doi.org/10.1016/0191-8869\(95\)00068-H](https://doi.org/10.1016/0191-8869(95)00068-H)
- Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, 8(2), 75–84.
<https://doi.org/10.1037//1076-898x.8.2.75>
- Martinussen, L. M., Møller, M., & Prato, C. G. (2017). Accuracy of young male drivers' self-assessments of driving skill. *Transportation Research Part F: Traffic Psychology and Behaviour*, 46, 228–235.
<https://doi.org/10.1016/j.trf.2017.03.001>
- Matthews, M. L., & Moran, A. R. (1986). Age differences in male drivers' perception of accident risk: The role of perceived driving ability. *Accident Analysis & Prevention*, 18(4), 299–313.
[https://doi.org/10.1016/0001-4575\(86\)90044-8](https://doi.org/10.1016/0001-4575(86)90044-8)

- McKenna, F. P., & Horswill, M. S. (2006). Risk taking from the participant's perspective: The case of driving and accident risk. *Health Psychology*, 25(2), 163–170. <https://doi.org/10.1037/0278-6133.25.2.163>
- McKnight, A. J., & McKnight, A. S. (2003). Young novice drivers: Careless or clueless? *Accident Analysis & Prevention*, 35(6), 921–925. [https://doi.org/10.1016/S0001-4575\(02\)00100-8](https://doi.org/10.1016/S0001-4575(02)00100-8)
- Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: Knowing about Knowing*. MIT Press.
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502–517. <https://doi.org/10.1037/0033-295X.115.2.502>
- Moran, C., Bennett, J. M., & Prabhakaran, P. (2019). Road user hazard perception tests: A systematic review of current methodologies. *Accident Analysis & Prevention*, 129, 309–333. <https://doi.org/10.1016/j.aap.2019.05.021>
- Murphy, P. R., Robertson, I. H., Harty, S., & O'Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. *eLife*, 4, e11946. <https://doi.org/10.7554/eLife.11946>
- Näätänen, R., & Summala, H. (1974). A model for the role of motivational factors in drivers' decision-making*. *Accident Analysis & Prevention*, 6(3), 243–261. [https://doi.org/10.1016/0001-4575\(74\)90003-7](https://doi.org/10.1016/0001-4575(74)90003-7)
- Norman, E., Pfuhl, G., Sæle, R. G., Svartdal, F., Låg, T., & Dahl, T. I. (2019). Metacognition in psychology. *Review of General Psychology*, 23(4), 403–424. <https://doi.org/10.1177/1089268019883821>
- Owsley, C., & McGwin, G. (2010). Vision and driving. *Vision Research*, 50(23),

2348–2361. <https://doi.org/10.1016/j.visres.2010.05.021>

Pluess, M. (2015). Individual Differences in Environmental Sensitivity. *Child Development Perspectives*, 9(3), 138–143. <https://doi.org/10.1111/cdep.12120>

Pluess, M., Assary, E., Lionetti, F., Lester, K. J., Krapohl, E., Aron, E. N., & Aron, A. (2018). Environmental sensitivity in children: Development of the Highly Sensitive Child Scale and identification of sensitivity groups. *Developmental Psychology*, 54(1), 51–70. <https://doi.org/10.1037/dev0000406>

Pluess, M., Lionetti, F., Aron, E. N., & Aron, A. (2023). People differ in their sensitivity to the environment: An integrated theory, measurement and empirical evidence. *Journal of Research in Personality*, 104, 104377. <https://doi.org/10.1016/j.jrp.2023.104377>

Pollatsek, A., Narayanaan, V., Pradhan, A., & Fisher, D. L. (2006). Using Eye Movements to Evaluate a PC-Based Risk Awareness and Perception Training Program on a Driving Simulator. *Human Factors*, 48(3), 447–464. <https://doi.org/10.1518/001872006778606787>

Rhodes, M. G. (2019). Metacognition. *Teaching of Psychology*, 46(2), 168–175. <https://doi.org/10.1177/0098628319834381>

Rubaltelli, E., Scrimin, S., Moscardino, U., Priolo, G., & Buodo, G. (2018). Media exposure to terrorism and people's risk perception: The role of environmental sensitivity and psychophysiological response to stress. *British Journal of Psychology*, 109(4), 656–673. <https://doi.org/10.1111/bjop.12292>

Sagberg, F., & Bjørnskau, T. (2006). Hazard perception and driving experience among novice drivers. *Accident Analysis & Prevention*, 38(2), 407–414. <https://doi.org/10.1016/j.aap.2005.10.014>

- Saraç, S., & Karakelle, S. (2012). On-line and off-line assessment of metacognition. *International Electronic Journal of Elementary Education*, 4(2), 301–315. <https://iejee.com/index.php/IEJEE/article/view/201>
- Scialfa, C. T., Borkenhagen, D., Lyon, J., Deschênes, M., Horswill, M., & Wetton, M. (2012). The effects of driving experience on responses to a static hazard perception test. *Accident Analysis & Prevention*, 45, 547–553. <https://doi.org/10.1016/j.aap.2011.09.005>
- Song, J., Kosovicheva, A., & Wolfe, B. (2024). Road Hazard Stimuli: Annotated naturalistic road videos for studying hazard detection and scene perception. *Behavior Research Methods*, 56(4), 4188–4204. <https://doi.org/10.3758/s13428-023-02299-8>
- Song, J., & Wolfe, B. (2024). Highly dangerous road hazards are not immune from the low prevalence effect. *Cognitive Research: Principles and Implications*, 9(1), 6. <https://doi.org/10.1186/s41235-024-00531-3>
- Sümer, N., Özkan, T., & Lajunen, T. (2006). Asymmetric relationship between driving and safety skills. *Accident Analysis & Prevention*, 38(4), 703–711. <https://doi.org/10.1016/j.aap.2005.12.016>
- Summala, H. (1988). Risk control is not risk adjustment: The zero-risk theory of driver behaviour and its implications. *Ergonomics*, 31(4), 491–506. <https://doi.org/10.1080/00140138808966694>
- Sun, L., Cheng, L., & Zhang, Q. (2021). The differences in hazard response time and driving styles of violation-involved and violation-free taxi drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 82, 178–186. <https://doi.org/10.1016/j.trf.2021.08.011>
- Sun, L., Hu, W., Cheng, L., & Zhang, C. (2024). Effects of hazard type and

- confidence level on hazard perception in young male drivers: An ERP study. *NeuroReport*, 35(5), 299. <https://doi.org/10.1097/WNR.0000000000002007>
- Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47(2), 143–148. [https://doi.org/10.1016/0001-6918\(81\)90005-6](https://doi.org/10.1016/0001-6918(81)90005-6)
- Vlakveld, W. P. (2014). A comparative study of two desktop hazard perception tasks suitable for mass testing in which scores are not based on response latencies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 22, 218–231. <https://doi.org/10.1016/j.trf.2013.12.013>
- Wang, J., Li, J., Xiao, Y., & Song, K. (2025). Steering in the dark: The impact of environmental luminance on driver behavior through optical flow analysis. *Attention, Perception, & Psychophysics*. <https://doi.org/10.3758/s13414-025-03146-y>
- Warren, W. H. (2006). The dynamics of perception and action. *Psychological Review*, 113(2), 358–389. <https://doi.org/10.1037/0033-295X.113.2.358>
- Wetton, M. A., Hill, A., & Horswill, M. S. (2011). The development and validation of a hazard perception test for use in driver licensing. *Accident Analysis & Prevention*, 43(5), 1759–1770. <https://doi.org/10.1016/j.aap.2011.04.007>
- White, M. J., Cunningham, L. C., & Titchener, K. (2011). Young drivers' optimism bias for accident risk and driving skill: Accountability and insight experience manipulations. *Accident Analysis & Prevention*, 43(4), 1309–1315. <https://doi.org/10.1016/j.aap.2011.01.013>
- Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1310–1321.

<https://doi.org/10.1098/rstb.2011.0416>

Zeuwts, L. H. R. H., Vansteenkiste, P., Deconinck, F. J. A., Cardon, G., & Lenoir, M. (2017). Hazard perception in young cyclists and adult cyclists. *Accident Analysis & Prevention*, 105, 64–71. <https://doi.org/10.1016/j.aap.2016.04.034>

4

Confidence and Risk in Hazard Perception and Dynamic Decision Making in Traffic

A study by

Lupia^{1,2*} E., Bontje^{1,3*} F., Bortolotti² A., Palumbo² R., and Janssen¹ C.P. (2025)

¹ Experimental Psychology & Helmholtz Institute, Utrecht University, Heidelberglaan 1, 3584 CS, Utrecht, The Netherlands.

² Department of Neuroscience, Imaging and Clinical Sciences, University “G.D'Annunzio” of Chieti-Pescara, Via Luigi Polacchi, 11, 66100 CH Chieti Scalo, Italy.

³ Sustainable Urban Mobility and Safety, Dutch Applied-Science Organization (TNO), Anna van Buerenplein 1, 2595 DA, The Hague, The Netherlands.

Abstract

Hazard perception is a widely used predictor of driving safety, yet perception alone does not ensure effective responses in dynamic traffic settings. Effective responses also rely on metacognitive judgments for decision making, such as confidence judgments and risk estimation. However, it is unknown how they manifest across different mobility modes. To investigate confidence, risk estimation and performance across perception (Driving Hazard Perception Task) and decision-making (Cycling Decision-Making Task) processes, within different mobility modes and environments. Forty-three participants completed a Hazard Perception Task and a Decision-Making Task. Trial-by-trial confidence and risk estimates were collected, alongside Hazard Perception Questionnaire (HPQ-6) scores, Environmental Sensitivity (HSP-12 scale) and self-reported general confidence. Confidence judgments showed significant positive cross-task correlations ($r = .71$). Risk estimates correlated significantly positively across tasks, particularly in hazardous conditions ($r = .45$). HPQ scores predicted confidence in hazardous trials, but not decision outcomes. Both tasks were also impacted by environmental factors. No direct relationship emerged between hazard perception accuracy and braking behavior, suggesting that perceptual skill does not fully translate to task performance. Confidence and risk estimation exhibit consistent patterns across mobility modes, but environmental and task-specific factors can modulate these relationships. While hazard perception tasks remain central to driver assessment, the use of confidence calibration provides a richer picture of self-monitoring and decision-making in traffic.

4.1 Introduction

Traffic safety remains a major societal concern, with human factors contributing to over 90% of fatal accidents (e.g., European Commission, 2024; Dingus et al., 2016). Part of the driving task involves making judgments, for example whether to yield for a pedestrian that is about to cross. When these and other judgments are poorly calibrated with actual skills and situational demands, drivers may overestimate their ability to handle hazards or underestimate the dangers they face (Horrey et al., 2015).

The calibration of these judgments is central to situational awareness (SA): the ability to perceive, understand, and anticipate the state of a dynamic environment (Endsley, 2020; Horrey et al., 2015). In traffic, poor SA can impair hazard perception and decision-making, reducing anticipation of future events, and increase the likelihood of adverse outcomes (Fisher & Strayer, 2014; Horswill, 2016; Endsley, 2020). According to Endsley's (2020) model, SA consists of three levels: perceiving environmental elements (Level 1), understanding their meaning (Level 2), and projecting their future state (Level 3).

Hazard perception has been primarily linked to the perceptual stage of SA (Level 1), as it requires the detection and encoding of salient environmental cues (e.g., Gugliotta et al., 2017; Katrahmani et al., 2017). By contrast, traffic decision-making – such as deciding whether to stop or continue cycling when approaching a crossing pedestrian – is better understood as a distinct process that closely draws on all three levels of SA (e.g., Endsley, 2020; Gugliotta et al., 2017).

Accurate hazard perception and dynamic decision making may also benefit from correctly calibrated metacognitive processes, such as confidence judgments and risk estimation. Although confidence judgments have been studied in high-risk domains such as aviation and the military (e.g., Sulistyawati & Chui, 2009; Sulistyawati et al., 2011), and in driving (e.g., Wohleber & Matthews, 2016; Sun et al., 2024; Hu et al., 2025; Lupia et al., 2025 submitted), the metacognitive consistency across different cognitive domains in mobility remain largely underexplored. Moreover, a recent meta-review of Hazard Perception Task suggests that while much previous research has examined how hazard perception is influenced by expertise, psychological

characteristics, and environmental factors (Cao et al., 2022), less attention has been paid to the effect of confidence. This manuscript therefore investigates performance and subjective assessment of confidence and risk in two tasks: Hazard Perception (in automotive scenarios) and Decision-Making (in cycling scenarios). This allowed us to assess how performance and subjective assessment correlate across different tasks and settings.

4.1.1 Perception and Decision Making in Traffic

Perception plays a fundamental role in traffic behavior (Castro, 2008), especially in safety-critical scenarios. Here, drivers need to perceive potential threats, anticipate hazards, and regulate their behavior for dynamic situations. The Hazard Perception Task is a well-established predictor of driving safety in this regard, with higher ability linked to lower crash risk (McKenna & Crick, 1994; Horswill & McKenna, 2004). Here, Hazard Perception (HP) refers to the skills or capabilities involved in detecting and identifying hazards (Borowsky et al., 2010).

Various factors including cognitive and psychological factors (e.g., experience, fatigue), substance use, and environmental factors (e.g., road type, lighting) influence the hazard perception ability (see meta-review by Cao et al., 2022). Important for the current study, the metacognitive calibration, can also affect performance on hazard perception (Sun et al., 2024; Hu et al., 2025; Lupia et al., 2025 submitted). In research, hazard perception has been tested through a variety of experimental paradigms (Cao et al., 2022; Moran et al., 2019). The most common approach is to use video-based HP tasks, in which participants respond when they perceive a developing hazard that requires immediate action (Wetton et al., 2011; Zeuwts et al., 2017; Sun et al., 2021; Crundall et al., 2021).

Beyond the ability to perceive hazards, safe driving also depends on how these perceptual inputs are translated into decisions (Zhang et al., 2017). Perceiving risk and acting upon it are distinct yet closely related processes, with decision quality strongly dependent on perceptual accuracy (Gianfranchi et al., 2021; Zhang et al., 2024). However, a gap remains in the literature about the role of metacognitive processes, such

as confidence and risk evaluation, in shaping perception and decision-making in mobility, especially regarding their consistency across domains.

4.1.2 The Role of Metacognition in Traffic

Confidence in decision-making refers to the metacognitive judgment of the probability that one's decision is correct (Fetsch et al., 2014; Kepecs & Mainen, 2012; Yeung & Summerfield, 2012). Confidence influences future decisions (Folke et al., 2017) and learning from feedback (Fleming & Daw, 2017). Two types of confidence can be distinguished: (1) local confidence, tied to a specific decision, and (2) global confidence, reflecting aggregated self-assessment over time (Rouault et al., 2019; Händel et al., 2020).

Local confidence generally increases when the evidence supporting a choice is stronger, and tends to decrease as response time increases (Kiani et al., 2014; Bontje & Zgonnikov, 2024). These effects are robustly demonstrated in simplified, non-naturalistic experimental paradigms (e.g. De Martino et al., 2013; Rouault et al., 2019). Bontje & Zgonnikov (2024) showed that these relationships also hold in more naturalistic traffic scenarios, where confidence is additionally shaped by kinematic factors such as time and distance gaps. Consistent with findings from lane-change and cooperative driving studies, smaller gaps were associated with greater decision uncertainty (Yan et al., 2015, 2023; Miller et al., 2022). While these studies did not measure confidence judgments directly, their results align with the notion that tighter kinematic constraints reduce confidence and increase perceived risk.

4.1.3 Study Aims and Hypotheses

Hazard perception (McKenna & Crick, 1994; Horswill & McKenna, 2004) and traffic decision-making (Yan et al., 2015, 2023; Bontje & Zgonnikov, 2024; Bontje et al., 2025, submitted) have been studied, yet their relationship remains unclear. In particular, little is known about how confidence and risk estimation vary and interrelate across different mobility modes and environmental conditions. To address this gap, a Driving Hazard Perception Task (Lupia et al., 2025 submitted) is combined with a Cycling Decision-Making Task (Bontje et al., 2025, submitted). The combination allows the

direct assessment and comparison of metacognitive consistency across perception (Hazard Perception Task) and decision-making (Decision-Making Task) contexts.

It is hypothesized that there is a significant positive correlation in confidence judgments and risk estimation across tasks, reflecting a trait-like component and metacognitive consistency. Conversely, if confidence and risk are adjusted to specific tasks, no correlation is hypothesized. Moreover, it is hypothesized that general hazard perception skill (assessed through a HPQ questionnaire) predicts higher task accuracy and better alignment of risks with actual hazards (cf. McKenna & Crick, 1994; Horswill & McKenna, 2004). Additionally, it is hypothesized that lower environmental sensitivity (Belsky & Pluess, 2009; Pluess, 2015), measured through the Highly Sensitive Person Scale (Pluess et al., 2023), will exhibit more stable performance under threatening conditions, as higher sensitivity has been linked to stronger stress reactivity (Aron et al., 2012; Rubaltelli et al., 2018) and higher risk perception (Lupia et al., 2025, submitted).

In the Hazard Perception Task, accuracy is expected to decline under nighttime conditions (cf. Asadamraji et al., 2019). In the Decision-Making Task, shorter time-gaps and higher pedestrian density are expected to result in more brake decisions and to shorter response times (cf. Bontje et al., 2025, submitted). Additionally, it is expected that confidence decreases under shorter time-gaps (i.e., TTAs, cf. Bontje & Zgonnikov, 2024).

4.2 Method

4.2.1 Participants

Forty-three participants took part (30 female; 11 male; 2 non-binary; age range 19-31 years, $M = 23.6$ years, $SD = 2.7$ years; 69.77% Dutch). Participants had a car driver's license for 1-12 years ($M = 4.60$ years; $SD = 3.01$ years). Utrecht is a cyclist-oriented city; accordingly, 65.1% of participants reported cycling almost daily, and all but one had learned to ride a bike before the age of nine. This research complied with the American Psychological Association Code of Ethics and was approved by the Ethical Review Board of the faculty of Social and Behavioral Sciences at Utrecht University

(code: 25-0119). Informed consent was obtained from each participant. They were compensated with course credit or 10 euro.

4.2.2 Materials

Self-Report Questionnaires

Four questionnaires were used: (1) a demographic questionnaire that also captured driving and cycling experience; (2) A standardized Hazard Perception Skill (HPQ-6; White et al., 2011) where participants rated their ability to perceive hazards compared to the average driver; (3) The standardized Highly Sensitive Person scale (HSP-12; Pluess et al., 2023), which captures self-reported sensitivity to external stimuli (i.e. environmental sensitivity); (4) a three-item questionnaire assessing general confidence in task performance, confidence in traffic participation, and general self-confidence. Details are available in Supplementary Materials (Sections A3, A4).

Balloon Analogue Risk Task

The Balloon Analogue Risk Task (BART; Lejuez et al., 2002) was included for comparison with another study, but not the main aim of the current paper (see Supplementary Materials, B4).

Hazard Perception Task

The (Driving) Hazard Perception Task, in E-Prime 3 (Psychology Software Tools, 2016; <https://pstnet.com/products/e-prime/>), was developed by Lupia and colleagues (2025, submitted). It consisted of 80 real-world traffic scenes from a first-person driver's perspective from the Road Stimulus Dataset (Song et al., 2024). Half of the videos contained a hazardous event and half did not; within each category, half were recorded during daytime and half at night. The hazardous scenarios were diverse (e.g., potential crashes with other vehicles, pedestrians, animals, and road obstacles), details are provided in Supplementary Materials (Section A2).

The task instructions started with a standard definition of a road hazard, used by Borowsky and colleagues (2010). Participants were instructed to respond to the

presence of a hazard by pressing the spacebar. Each clip terminated immediately upon the response, or otherwise at the end of the clip.

Decision-Making Task

The (Cycling) Decision-Making Task was developed by Bontje et al. 2025 (submitted), of which a subset of stimuli was used. Participants viewed video fragments that simulated movement through urban areas from the egocentric perspective of a cyclist. The traveling speed of the cyclist was determined by predefined time-to-arrival values. Pedestrian density varied between conditions (low/high), with the high-density condition containing approximately four times as many pedestrians as the low-density condition. Pedestrians moved at 1.4 m/s on average ($SD = 0.1$ m/s) and walked parallel to or waiting alongside the cycling lane. Two city-center locations (Stadhuisplein, Janskerkhof; mimicking areas in Utrecht city) and four unique routes (two per location, opposing directions) were used.

On each trial, one pedestrian crossed and paused at a location ahead of the participant's position. Participants had to respond to this event as fast as possible: Continue cycling (press 'up' arrow key) or Stop/Brake (down' arrow key). After a key press, the video terminated.

Confidence Judgment and Subjective Risk Estimation: Trial-by-Trial Ratings

In each task, after each trial, participants first rated confidence in task performance (cf. Sun et al., 2024) and then perceived risk (cf. Castro et al., 2020). Figures 4.1 and 4.2 illustrate a complete trial of the Hazard Perception Task and the Decision-Making Task.

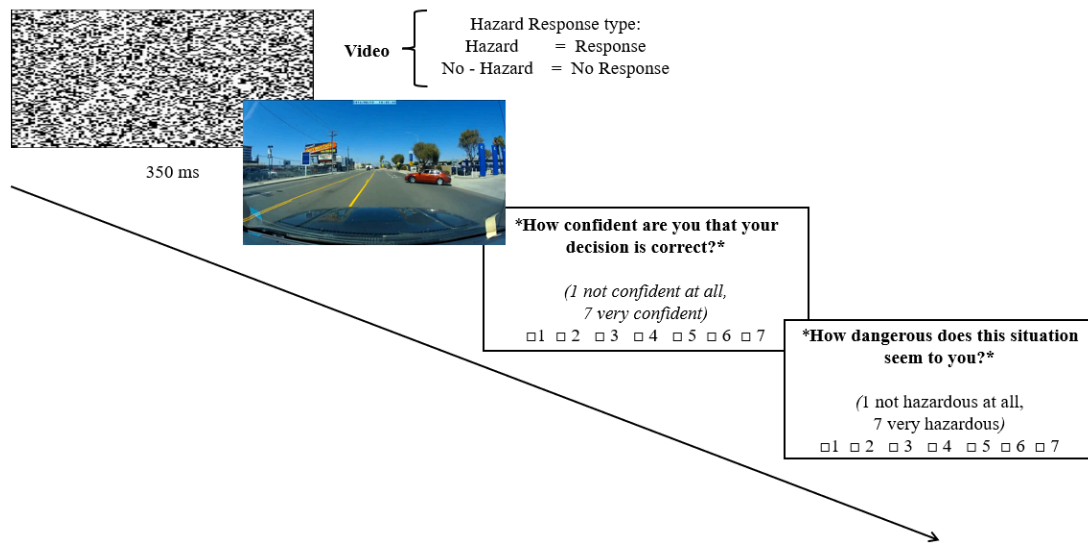


Figure 4.1 *Sequence of events in a trial of the Hazard Perception Task.*

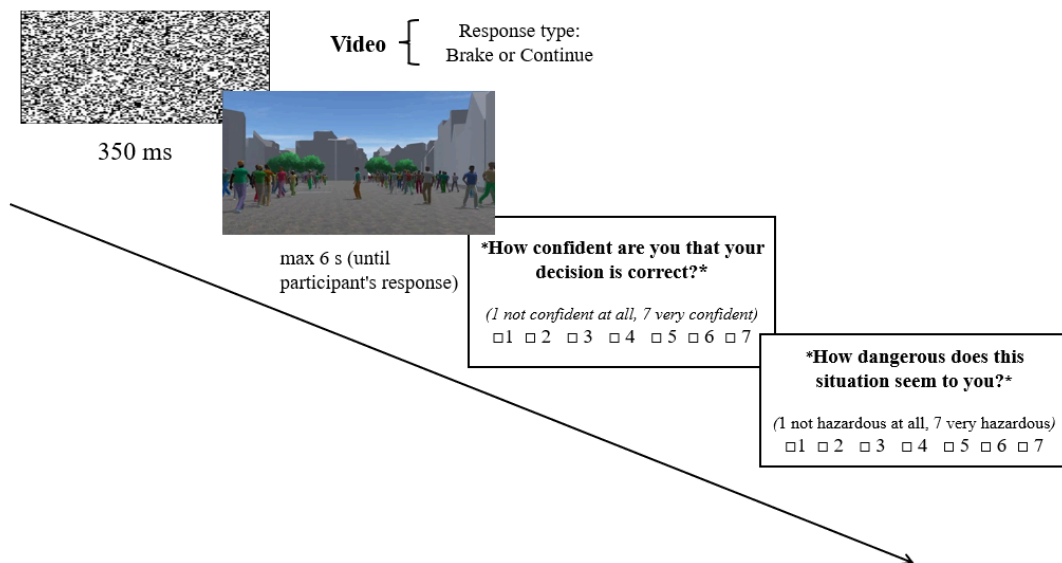


Figure 4.2 *Sequence of events in a trial of the Decision-Making Task.*

4.2.3 Design

Hazard Perception Task

In the (Driving) Hazard Perception Task, hazard presence (hazardous vs. non-hazardous) and lighting conditions (daytime vs. nighttime) were manipulated within-subjects (see Figure 4.3 for examples). There were 80 trials (20 trials per condition) in a random order.



Figure 4.3 Example frames from the Hazard Perception Task, showing clips containing hazards (A, C) and without hazardous events (B, D), presented during daytime (A, B) and nighttime (C, D) conditions.

Decision-Making Task

In the (Cycling) Decision-Making Task, pedestrian density (low vs. high) and time-to-arrival (TTA; 4 seconds vs. 6 seconds, corresponding to speeds of 13.5 km/h and 9 km/h) were manipulated within-subjects. Simulations were made at two video locations that had two cycling directions of traveling, resulting in 16 unique videos ($2 \times 2 \times 2 \times 2$) (see Figure 4.4 for examples). Each video was shown six times. Trials were grouped in two blocks of 48 trials with a five-minute break between blocks. Within each block, all 16 unique videos were shown three times in a randomized order with the constraint that no two consecutive trials/videos were identical. For counterbalancing purposes, for each randomized sequence that was assigned to one participant, another participant received the inverse order.



Figure 4.4 Example screenshots from scenarios of the Decision-Making Task in areas with high (A, C) and low density (B, D) and representing different environments (A, B at the Janskerkhof, and C, D at the Stadhuisplein).

4.2.4 Procedure

Participants received a detailed explanation of the study and could ask questions before providing written informed consent. They first completed three questionnaires: (1) demographics (2) HPQ-6, and (3) HSP-12 scale. Next, participants completed the Balloon Analogue Risk task, followed by the main tasks: the Hazard Perception Task and then the Decision-Making Task. Finally, participants again completed the HPQ-6, followed by the three general confidence questions. The full session lasted approximately one hour.

4.2.5 (Statistical) Analysis

Hazard Perception Task

For the hazard clips, response time is the difference between the moment the participant pressed the space button and the annotated hazard onset time (defined by Song et al., 2024). Trials were excluded if responses (1) were within 0.1 s of hazard onset (8.5 % of trials), (2) occurred after the annotated driver response time (i.e., the pre-defined moment when an evasive maneuver—such as braking or swerving – was visibly initiated in the video, 14 %), (3) had a confidence value outside of the 7-point Likert scale options (i.e., another key on the keyboard was pressed; 0.037 %, i.e. one trial), and (4) had a subjective risk value outside of the 7-point Likert scale options (0 %). The percentage of valid trials per participant was between 70.0% and 87.5% ($M=77.6\%$; $SD=4.22\%$).

Response accuracy was scored binary: hits (i.e., detection of hazard) and correct rejections (i.e., no hazard) were marked as ‘correct’; false alarms and misses were marked as ‘incorrect’.

Generalized linear mixed-effects models (GLMMs) were used to examine the relationships between the dependent variables (accuracy, response time, risk estimation judgment, and local confidence judgment) and independent variables (hazard presence, time of day [nighttime/daytime], HSP score, HPQ score, previous local confidence judgment, general driving confidence, and general life confidence). Random intercepts accounted for between-participant variability. The *day* and the *no hazard* conditions

were defined as the reference categories. Model selection was based on the Bayesian Information Criterion (BIC). All analyses were performed in MATLAB R2023a using the *fitglme* function. Results are reported relative to the baseline condition (i.e.: non-hazardous, daytime, and all continuous variables = 0).

Decision-Making Task

Response time was defined (by Bontje et al., 2025, submitted) relative to the first step that the pedestrian made perpendicular to the path of the “cyclist”. The following trials were excluded: (1) trials with no response (3.8% of trials), (2) responses made before the appearance of the crossing pedestrian (5.0%), (3) responses made after passing the crossing pedestrian (0.28%), (4) confidence judgments with values outside of the 7-point Likert scale options (0 %), and (5) subjective risk estimates with values outside the 7-point Likert scale options (0 %). The percentage of valid trials per participant was between 54.7% and 100% (M = 91.8% ; SD =10.9%).

The analysis mirrored that used in the hazard perception task. BIC-selected generalized linear mixed-effects models (GLMMs) were employed to examine the relationships between dependent (decision outcome, response time, risk estimation, local confidence judgment) and independent variables (TTA, pedestrian density [high/low], HSP score, HPQ score, previous local confidence judgment, general driving confidence, and general life confidence). Random intercepts accounted for between-participant variability and differences due to location. Results are reported relative to the baseline condition (i.e.: high pedestrian density condition, continue decision, and all continuous variables = 0).

Correlation Analyses between Hazard Perception and Decision-Making

Five sets of correlation analyses were conducted:

1. Accuracy: The correlation between participants’ mean accuracy in the (driving) hazard perception task and their mean percentage of braking decisions in the (cycling) decision making task.

2. Risk estimates: Three correlations between mean risk estimates in the two tasks: overall, and for hazardous and non-hazardous (driving) trials.
3. Confidence judgments: The correlation between participants' mean local confidence judgments in the two tasks: overall, and for hazardous and non-hazardous (driving) trials.
4. Hazard Perception Questionnaire (HPQs): The correlation between HPQs before and after the experiment.
5. General confidence: The correlations between participants' self-reported general confidence: in task execution, in driving skills, and in everyday life.

All materials, including the data and video stimuli used for collection and analysis, are available in the OSF repository at

https://osf.io/gxbpe/overview?view_only=d2e12ef2509e4bc8871d0f75d208c98c

4.3 Results

4.3.1 Hazard Perception Task

Table 4.1 summarizes the results of the (G)LMM analysis. The main text only discusses the main significant effects. Full model specifications, including non-significant or negligible (i.e., a percentage point change smaller than 0.01 or a relative percentage change smaller than 0.01%) are provided in Supplementary Material (Section B5-B10).

Overall performance: Participants detected hazards with high accuracy (92%). In the baseline (non-hazardous daytime trial, continuous predictors set to zero), accuracy was estimated at 100%. Accuracy dropped sharply in hazardous scenarios (−59.31 pp, $p < .001$) and was slightly lower in nighttime scenarios (−0.23 pp, $p = .021$).

Response times: Baseline response time was 0.76 s. Responses were faster in nighttime scenarios (−0.09 s, $p = .003$).

Risk estimates: Baseline risk score was 3.68 (Likert) points. Risk was higher in nighttime scenarios (+0.37 point, $p = .030$), lower after correct responses (−1.51 point, $p < .001$), and reduced by higher HPQ scores (−0.22 point, $p = .048$). An interaction effect between hazard and accuracy indicated that in hazardous trials, higher accuracy predicted higher risk estimates (+3.57 point, $p < .001$). In hazardous-only trials (baseline = 5.53 point), an interaction between nighttime and response time showed that faster nighttime scenarios responses predicted lower risk (−0.38 point per second, $p = .038$).

Confidence judgments: Baseline confidence was 2.72 (Likert) point. Confidence was higher after correct responses (+3.37 point, $p < .001$), in nighttime scenarios (+1.45 point, $p < .001$), in hazardous trials (+0.89 point, $p = .038$), with higher subjective risk (+0.48 point, $p < .001$), and following previous confidence judgments (+0.05 point, $p = .006$). Several interactions moderated these effects: risk contributed less to confidence during nighttime scenarios (−0.40 point, $p < .001$), in hazardous trials (−0.42 point, $p < .001$), and after correct responses (−0.87 point, $p < .001$). Accuracy effects were also weaker during nighttime video (−1.70 point, $p < .001$) and in hazardous trials (−2.38 points, $p < .001$). These reductions by accuracy were partly offset by compensatory effects of risk estimation (+1.00 point in hazardous, +0.45 point during nighttime scenarios; both $p < .001$). In hazardous-only trials, confidence increased with HPQ (+1.07 point, $p < .001$), risk estimation (+0.87 point, $p < .001$), and previous confidence (+0.05 point, $p = .024$), but risk effects weakened with higher HPQ (−0.15 point, $p < .001$).

Predictor	Accuracy (pp)	RT (s)	Risk all (points)	Risk hazard (points)	Confidence all (points)	Confiden ce hazard (points)
Nighttime	-0.23*	-0.09**	+0.37*	ns	+1.45***	ns
Hazard	-59.31***	—	ns	—	+0.89*	—
Accuracy	—	—	-1.51***	—	+3.37***	—
RT	—	—	—	ns	—	ns
HSP	-0.01**	ns	ns	ns	ns	ns
HPQ	ns	ns	-0.22*	ns	ns	+1.07***
Risk estimation	—	—	—	—	+0.48***	+0.87***
Prev. confidence	ns	ns	ns	ns	+0.05**	+0.05*
General confidence	ns	ns	ns	ns	ns	ns
Key interactions	Hazard×HS P: +0.002***	—	Hazard×Accur acy: +3.57***	Nighttime× RT: -0.38*	multiple (see text)	Risk×HP Q: -0.15***

Table 4.1 Predictors of hazard detection accuracy, response times, risk estimates, and confidence judgments, pp = percentage points; s = seconds; ns = non-significant. Significance: *: $p < .05$; **: $p < .01$; ***: $p < .001$.

4.3.2 Decision-Making Task

Table 4.2 summarizes the results of the (G)LMM analysis. The main text again only covers the main significant effects. Full model specifications are provided in Supplementary Material (Section B11-B14).

Braking behaviour: Participants braked in 84.1% of the trials. The proportion of braking in the baseline condition (i.e. high pedestrian density and all continuous variables = 0) was 79.7%. Braking was less likely in low pedestrian density (-8.94 pp, $p < .001$) and with larger time-to-arrival (TTA; -6.17 per 1-s increase, $p < .001$).

Response times: The baseline response time was -0.73 s. Brake decisions were faster than continue decisions (-1.07 s, $p < .001$). Moreover, response times increased with a larger TTA (+0.19 s per 1-s increase, $p < .001$), this was stronger in brake decisions (+0.29 s per 1-s decrease, $p < .001$).

Risk estimation: The baseline risk estimation was 3.51 points. Risk was lower in low-density environments (-0.92 point, $p < .001$), and increased with longer response times (+1.07 point per 1-s increase, $p = .003$), and when braking (+1.72 point, $p = .010$). Interactions showed longer response times increased risk in low-density (+0.29 point per 1-s, $p < .001$), whereas with large TTA larger response times decreased risk (-0.12 point per 1-s, $p = .017$).

Confidence judgments: The baseline confidence judgment was 6.18 points. Confidence was higher in low-density environments (+0.80 point, $p < .001$), and increased with response times (-0.66 point per 1-s, $p = .037$), risk (-0.58 point per risk point, $p < .001$), and when braking (-2.46 point, $p < .001$). Confidence increased slightly with previous confidence ratings (+0.06 point per previous rating point, $p < .001$). Interactions shaped these effects: confidence dropped with long response times and braking in low-density (-0.51 point per 1-s and -1.2 points, $ps < .001$), but increased when all three co-occurred (+0.67 point, $p < .001$). The negative response-time effect was reduced with large TTA (+0.16 point, $p = .002$) or braking (+0.92 point, $p = .005$), though with large TTA, long RT, and braking combined, confidence decreased (-0.21 point, $p < .001$). Finally, confidence increased when braking in high-risk estimated situations (+0.73 point per point of risk, $p < .001$).

Predictor	Braking (pp)	Response time (s)	Risk estimate (points)	Confidence judgment (points)
Low density	-8.94***	ns	-0.92***	+0.80***
TTA	-6.17***	+0.19***	ns	ns
Brake decision	—	-1.07***	+1.72*	-2.46***
RT	—	—	+1.07**	-0.66*
Risk estimate	—	—	—	-0.38***
Previous confidence	ns	ns	ns	+0.06***
HSP	ns	ns	ns	ns
HPQ	ns	ns	ns	ns
General confidence	ns	ns	ns	ns
Key interactions	—	Brake×TTA: 0.38***	Low density×RT: +0.29*** TTA×RT: -0.12*	multiple (see text)

Table 4.2 Predictors of braking, response times, risk estimates, and confidence. Notes. *pp* = percentage point; *points* = raw scale points; *s* = seconds. *ns* = non-significant. Significance: *: $p < .05$; **: $p < .01$; ***: $p < .001$.

4.3.3 Correlation Analyses between Hazard Perception and Decision-Making

Performance: There was no significant correlation between accuracy in Hazard Perception Task and percentage of braking decisions in Decision-Making Task ($r = -.03$, $p = .87$, Figure 4.5A).

Risk estimation: A significant positive correlation emerged between participants' mean risk estimations in the two tasks ($r = .33$, $p = .03$, Figure 4.5B). When hazardous and non-hazardous driving trials were correlated separately with the

cycling task (see Figure 4.6A - 4.6B), hazard trials showed a significant positive correlation ($r = .45$, $p = .003$), whereas non-hazard trials showed no significant relationship ($p = .22$).

Confidence judgments: Participants' mean local confidence judgments across tasks showed a significant positive correlation ($r = .71$, $p < .001$, Figure 4.5C). This also held when hazardous ($r = .62$, $p < .001$) and non-hazardous trials were analyzed separately ($r = .63$, $p < .001$) (see Figure 4.6C - 4.6D).

Hazard Perception Questionnaires (HPQs): A significant positive correlation was found between HPQ score before and after the experiment ($r = .79$, $p < .001$).

General confidence: General confidence in task execution significantly correlated with general confidence in driving skills ($r = .47$, $p < .001$), and to a smaller degree with general confidence in everyday life ($r = .23$, $p < .001$). Additionally, a significant correlation was found between confidence in driving skills and confidence in everyday life ($r = .44$, $p < .001$).

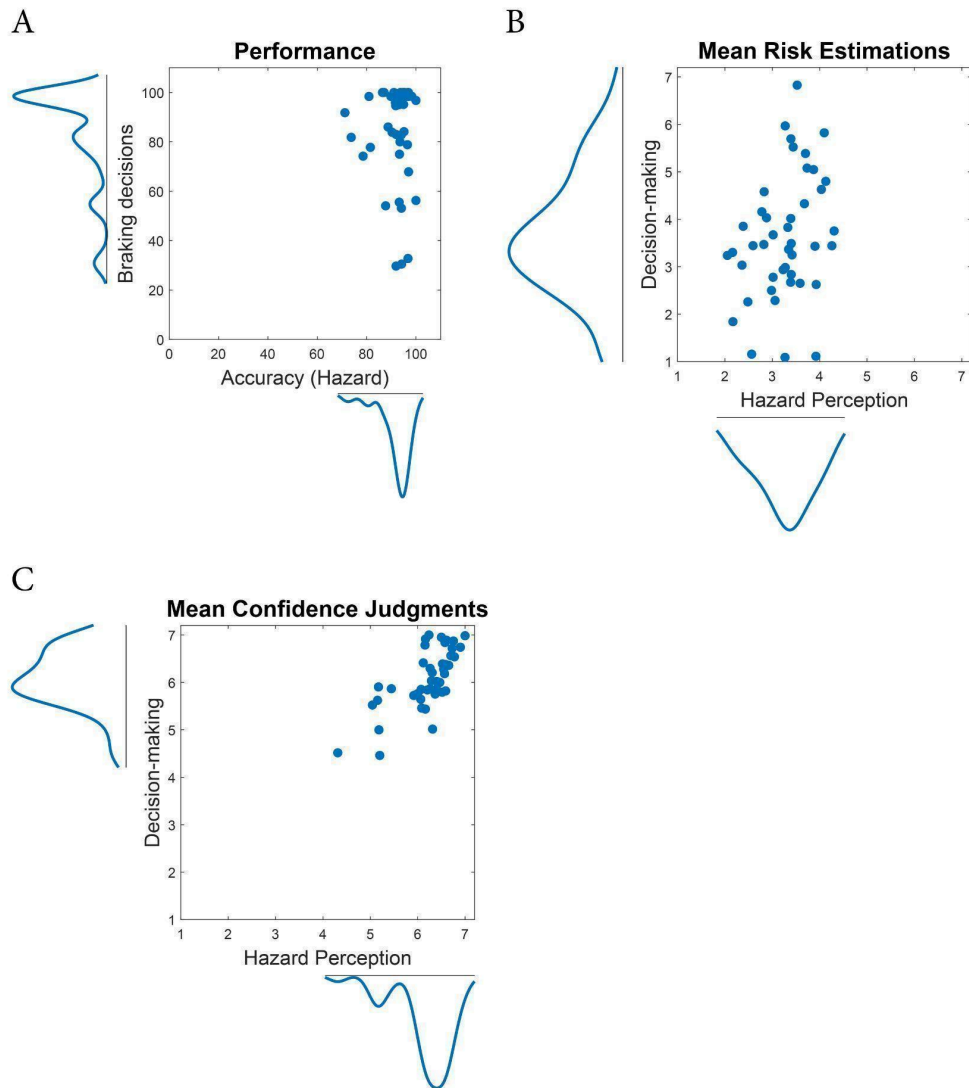


Figure 4.5. Marginal density plots display distributions between (Driving) Hazard Perception Task and (Cycling) Decision-Making Task. Panels show A) performance: accuracy vs. braking, B) mean risk estimation, and C) mean confidence judgments.

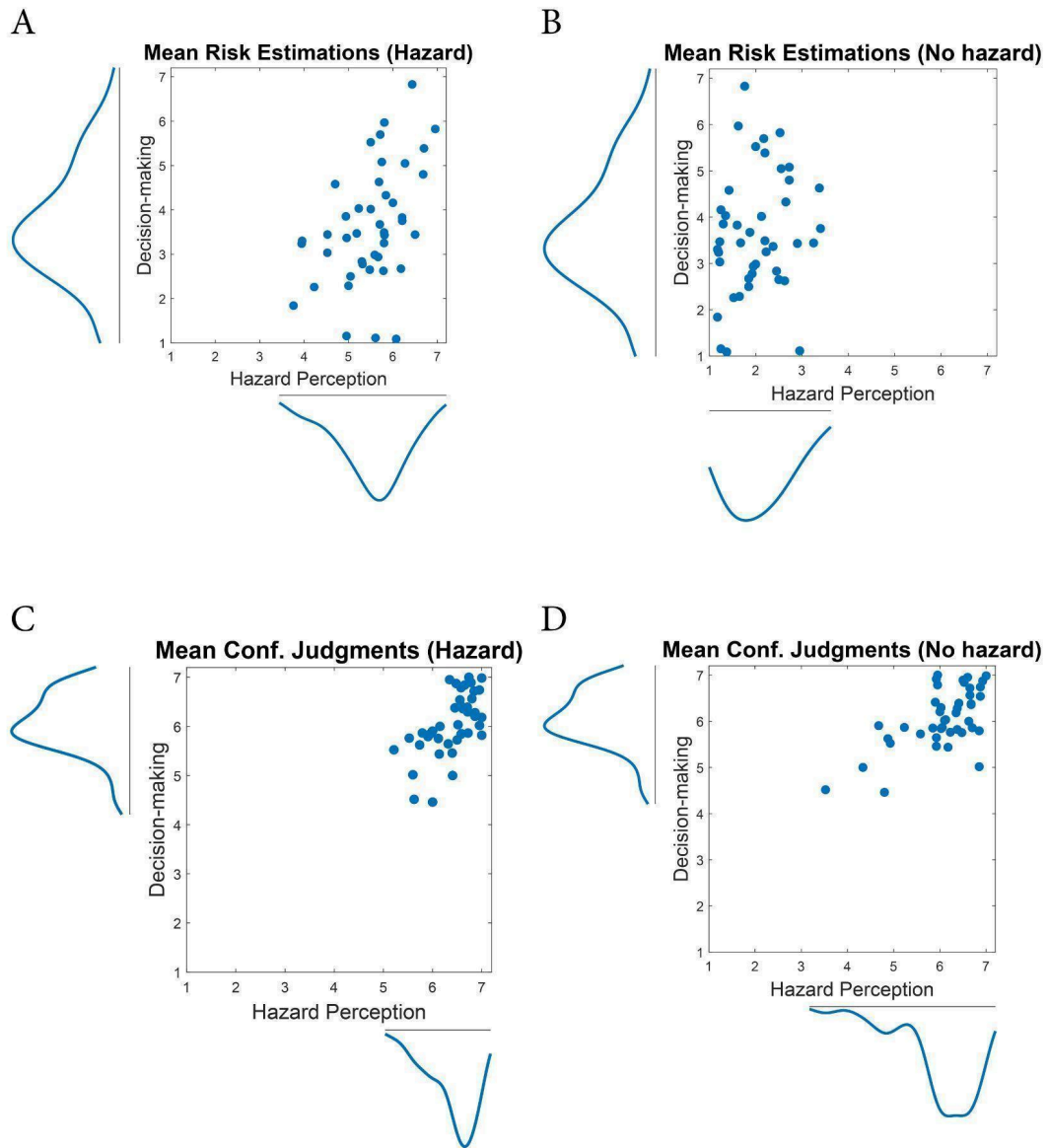


Figure 4.6. Marginal density plots display distributions between (Driving) Hazard Perception Task and (Cycling) Decision-Making Task. Panels show A) mean risk estimation in hazard trials, B) mean risk estimation in no hazard trials, C) mean confidence judgments in hazard trials, and D) mean confidence judgments in no hazard trials.

4.4 Discussion

The present paper investigated how risk estimation and confidence judgments relate across a Driving Hazard Perception Task and a Cycling Decision-Making Task. The results indicate cross-task consistency: a positive correlation for confidence judgments and a positive correlation for mean risk estimation (see Figure 4.5 - 4.6). This consistency across the tasks using different stimulus types suggests that it reflects a generalizable metacognitive tendency rather than stimulus-specific effects. The HPQ scores showed good pre–post stability, and the general confidence measures correlated moderately across domains. Confidence reflects objective task performance, as accuracy in the Driving Hazard Perception Task was a significant predictor of confidence judgments.

There was no significant correlation between performance on the two tasks (hazard detection accuracy vs. braking percentage), suggesting that these performance measures do not directly relate to each other. Task-specific analyses confirmed some expected effects while others did not emerge. In the Driving Hazard Perception Task, accuracy decreased under nighttime conditions (replicating earlier findings; Asadamraji et al., 2019; Horswill, 2016; Cao et al., 2022). Environmental sensitivity (HSP scale) had a significant, yet in practice very small effect on accuracy. Although consistent with literature on contextual cues (Belsky & Pluess, 2009; Aron et al., 2012; Pluess, 2015; Pluess et al., 2018) and threat awareness (Rubaltelli et al., 2018), its impact on hazard detection requires further exploration. The absence of the relationship between accuracy and HPQ scores or general confidence partially diverges from studies reporting links between self-reported hazard perception and task performance (e.g., Åbele et al., 2018). This suggests that subjective self-reported measures may not always predict actual performance in safety-traffic scenarios. Risk estimates proved to be context-sensitive, increasing during nighttime scenarios (cf. Evans et al., 2020) and decreasing after correct responses (cf. research that risk estimation varies with task performance and hazard type; Sun et al., 2021, 2024; Rubaltelli et al., 2018). Confidence was higher after correct responses, during nighttime condition, and in hazardous trials, though moderated by hazard presence and accuracy, echoing earlier findings on confidence bias in situation awareness (Sulistyawati et al., 2011) and overconfidence in driving (Wohleber

& Matthews, 2016). In the cycling task, shorter time-gaps increased braking likelihood and reduced response times (cf. Bontje et al., 2025, submitted; Bontje & Zgonnikov, 2024). Pedestrian density increased the odds of braking (cf. Bontje et al., 2025 submitted), elevated risk estimates, and lowered confidence, which was also in line with earlier work showing that traffic interactions heighten perceived risk (Zeuwts et al., 2017; European Commission, 2024).

4.4.1 Limitations and Implications

There were limitations to the study. First, data was collected in a controlled laboratory setting, where behavior had no real-life consequences or incentives, potentially reducing ecological validity. Future studies could address this by using field-based or semi-naturalistic designs. Second, is the high proportion of hazardous trials (50%) in the Driving Task, which may have influenced participants' expectations. Finally, in the Cycling Task participants did not physically cycle.

Additionally, the participants were recruited in the Netherlands, where cycling is a common practice supported by well-developed infrastructure. While this likely reduces behavioral variability, it may limit generalizability to countries with a less developed cycling culture. Future studies could test whether these patterns hold in other countries as well.

The findings have two practical implications for traffic safety. First, they suggest that confidence and risk estimation play a key role in perception and adaptive decision-making. Therefore, this could be an area of attention for future training programs aimed at improving traffic safety. Second, hazard perception tests that are used in licensing procedures (Wetton et al., 2011) could be extended with confidence judgments or risk estimates, to obtain a richer assessment of self-monitoring skills, as it could help identify individuals who systematically misjudge their abilities or underestimate risks (i.e., that are not well calibrated). Similarly, screening for hazard perception skill (HPQ) may help identify individuals who would benefit most from training aimed at improving confidence–risk calibration.

4.5 Conclusion

This study demonstrates that confidence judgments and risk estimation generalize across different traffic contexts, whereas performance is more tightly bound to (task/test specific) situational demands such as nighttime (on driving task), or pedestrian density and time-gaps (on cycling task). Confidence and risk estimation thus appear partly trait-like and transferable across domains, while performance reflects immediate environmental constraints.

References

- Åbele, L., Haustein, S., Møller, M., & Martinussen, L. M. (2018). Consistency between subjectively and objectively measured hazard perception skills among young male drivers. *Accident Analysis & Prevention*, *118*, 214–220. <https://doi.org/10.1016/j.aap.2018.02.022>
- Aron, E. N., Aron, A., & Jagiellowicz, J. (2012). Sensory Processing Sensitivity: A Review in the Light of the Evolution of Biological Responsivity. *Personality and Social Psychology Review*, *16*(3), 262–282. <https://doi.org/10.1177/1088868311434213>
- Asadamraji, M., Saffarzadeh, M., Ross, V., Borujerdian, A., Ferdosi, T., & Sheikholeslami, S. (2019). A novel driver hazard perception sensitivity model based on drivers' characteristics: A simulator study. *Traffic Injury Prevention*, *20*(5), 492–497. <https://doi.org/10.1080/15389588.2019.1607971>
- Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin*, *135*(6), 885–908. <https://doi.org/10.1037/a0017376>
- Bontje, F., Spierings, B., Geraerts, R., Bakri, T., van Maanen, L., & Janssen, C. P. (2025). The influence of spatial and social context on cyclists' anticipatory yielding behavior. *[Manuscript Submitted for Publication]*. *Experimental Psychology & Helmholtz Institute, Utrecht University*.
- Bontje, F., & Zgonnikov, A. (2024). How Sure is the Driver? Modelling Drivers' Confidence in Left-Turn Gap Acceptance Decisions. *Computational Brain & Behavior*, *7*(3), 437–456. <https://doi.org/10.1007/s42113-024-00207-7>
- Borowsky, A., Oron-Gilad, T., & Parmet, Y. (2010). The role of driving experience in hazard perception and categorization: A traffic-scene paradigm. *World*

-
- Academy of Science, Engineering and Technology*, 66, 305–309.
<https://www.scopus.com/pages/publications/79951530926>
- Cao, S., Samuel, S., Murzello, Y., Ding, W., Zhang, X., & Niu, J. (2022). Hazard Perception in Driving: A Systematic Literature Review. *Transportation Research Record: Journal of the Transportation Research Board*, 2676(12), 666–690. <https://doi.org/10.1177/03611981221096666>
- Castro, C. (2008). *Human Factors of Visual and Cognitive Performance in Driving* (0 ed.). CRC Press. <https://doi.org/10.1201/9781420055337>
- Chan, E., Pradhan, A. K., Pollatsek, A., Knodler, M. A., & Fisher, D. L. (2010). Are driving simulators effective tools for evaluating novice drivers' hazard anticipation, speed management, and attention maintenance skills? *Transportation Research Part F: Traffic Psychology and Behaviour*, 13(5), 343–353. <https://doi.org/10.1016/j.trf.2010.04.001>
- Crundall, D., Van Loon, E., Baguley, T., & Kroll, V. (2021). A novel driving assessment combining hazard perception, hazard prediction and theory questions. *Accident Analysis & Prevention*, 149, 105847. <https://doi.org/10.1016/j.aap.2020.105847>
- De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, 16(1), 105–110. <https://doi.org/10.1038/nn.3279>
- Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, 113(10), 2636–2641. <https://doi.org/10.1073/pnas.1513271113>
- Endsley, M. (2020). Situation Awareness in Driving. In *Handbook of Human*

Factors for Automated, Connected, and Intelligent Vehicles. CRC Press.

E-Prime® Stimulus Presentation Software | Psychology Software Tools. (2017, September 11). Psychology Software Tools | Solutions for Research, Assessment, and Education. <https://pstnet.com/products/e-prime/>

European Commission. (2024). *Road safety thematic report – Main factors causing fatal crashes*. https://road-safety.transport.ec.europa.eu/document/download/a7428369-8eaf-4032-806e-ea08b46028c0_en?filename=ERSO-TR-MainCauses.pdf

Evans, T., Stuckey, R., & Macdonald, W. (2020). Young drivers' perceptions of risk and difficulty: Day versus night. *Accident Analysis & Prevention*, 147, 105753. <https://doi.org/10.1016/j.aap.2020.105753>

Fetsch, C. R., Kiani, R., & Shadlen, M. N. (2014). Predicting the Accuracy of a Decision: A Neural Mechanism of Confidence. *Cold Spring Harbor Symposia on Quantitative Biology*, 79, 185–197. <https://doi.org/10.1101/sqb.2014.79.024893>

Fisher, D. L., & Strayer, David. L. (2014). Modeling Situation Awareness and Crash Risk. *Annals of Advances in Automotive Medicine*, 58, 33–39. <https://pmc.ncbi.nlm.nih.gov/articles/PMC4001668/>

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, 124(1), 91–114. <https://doi.org/10.1037/rev0000045>

Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2016). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour*, 1(1), 0002. <https://doi.org/10.1038/s41562-016-0002>

- Gianfranchi, E., Tagliabue, M., Megías-Robles, A., & Maldonado, A. (2021). The role of feedback in shaping responses to risky road scenarios: Evidence from electrodermal activity. *Transportation Research Part F: Traffic Psychology and Behaviour*, 83, 130–147. <https://doi.org/10.1016/j.trf.2021.10.001>
- Gugliotta, A., Ventsislavova, P., Garcia-Fernandez, P., Peña-Suarez, E., Eisman, E., Crundall, D., & Castro, C. (2017). Are situation awareness and decision-making in driving totally conscious processes? Results of a hazard prediction task. *Transportation Research Part F: Traffic Psychology and Behaviour*, 44, 168–179. <https://doi.org/10.1016/j.trf.2016.11.005>
- Horrey, W. J., Lesch, M. F., Mitsopoulos-Rubens, E., & Lee, J. D. (2015). Calibration of skill and judgment in driving: Development of a conceptual framework and the implications for road safety. *Accident Analysis & Prevention*, 76, 25–33. <https://doi.org/10.1016/j.aap.2014.12.017>
- Horswill, M. S. (2016). Hazard Perception in Driving. *Current Directions in Psychological Science*, 25(6), 425–430. <https://doi.org/10.1177/0963721416663186>
- Horswill, M. S., & McKenna, F. P. (2004). Drivers' hazard perception ability: Situation awareness on the road. In S. Banbury & T. S. (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 155–175). Ashgate. <https://centaur.reading.ac.uk/14289/>
- Hu, W., Sun, L., & Cheng, L. (2025). Effects of confidence level and hazard type on the visual search patterns and hazard response times of young drivers. *Traffic Injury Prevention*, 1–7. <https://doi.org/10.1080/15389588.2025.2497521>
- Katrahmani, A., Ahmadi, N., & Romoser, M. (2017). Using situation awareness as a measure of driver hazard perception ability. In *Driving Assessment*

Conference (Vol. 9). Iowa City, IA: University of Iowa (pp 256-262).

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1322–1337. <https://doi.org/10.1098/rstb.2012.0037>

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice Certainty Is Informed by Both Evidence and Decision Time. *Neuron*, 84(6), 1329–1342. <https://doi.org/10.1016/j.neuron.2014.12.015>

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, 8(2), 75–84. <https://doi.org/10.1037/1076-898X.8.2.75>

Lupia, E., Bortolotti, A., & Palumbo, R. (2025). The Impact of Overconfidence and Environmental Conditions on Hazard Perception and Risk Assessment: An Experimental Study Using Video-Based Traffic Scenarios. *[Pre-print Submitted for Review]*. Department of Neuroscience, Imaging and Clinical Sciences, University “G.D’Annunzio” of Chieti-Pescara.

Malone, S., & Brünken, R. (2015). Hazard Perception Assessment – How Much Ecological Validity is Necessary? *Procedia Manufacturing*, 3, 2769–2776. <https://doi.org/10.1016/j.promfg.2015.07.713>

Malone, S., & Brünken, R. (2016). The role of ecological validity in hazard perception assessment. *Transportation Research Part F: Traffic Psychology and Behaviour*, 40, 91–103. <https://doi.org/10.1016/j.trf.2016.04.008>

McKenna, F. P., & Crick, J. L. (1994). Hazard perception in drivers: A

- methodology for testing and training. *TRL Contractor Report*, 313. <https://trid.trb.org/View/457386>
- Miller, L., Leitner, J., Kraus, J., Lee, J., Daimon, T., Kitazaki, S., & Baumann, M. (2022). Time to Arrival as Predictor for Uncertainty and Cooperative Driving Decisions in Highly Automated Driving. *2022 IEEE Intelligent Vehicles Symposium (IV)*, 1048–1053. <https://doi.org/10.1109/IV51971.2022.9827416>
- Moran, C., Bennett, J. M., & Prabhakaran, P. (2019). Road user hazard perception tests: A systematic review of current methodologies. *Accident; Analysis and Prevention*, 129, 309–333. <https://doi.org/10.1016/j.aap.2019.05.021>
- Pluess, M. (2015). Individual Differences in Environmental Sensitivity. *Child Development Perspectives*, 9(3), 138–143. <https://doi.org/10.1111/cdep.12120>
- Pluess, M., Assary, E., Lionetti, F., Lester, K. J., Krapohl, E., Aron, E. N., & Aron, A. (2018). Environmental sensitivity in children: Development of the Highly Sensitive Child Scale and identification of sensitivity groups. *Developmental Psychology*, 54(1), 51–70. <https://doi.org/10.1037/dev0000406>
- Pluess, M., Lionetti, F., Aron, E. N., & Aron, A. (2023). People differ in their sensitivity to the environment: An integrated theory, measurement and empirical evidence. *Journal of Research in Personality*, 104, 104377. <https://doi.org/10.1016/j.jrp.2023.104377>
- Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, 10(1), 1141. <https://doi.org/10.1038/s41467-019-09075-3>
- Rubaltelli, E., Scrimin, S., Moscardino, U., Priolo, G., & Buodo, G. (2018). Media exposure to terrorism and people's risk perception: The role of environmental sensitivity and psychophysiological response to stress. *British Journal of*

Psychology, 109(4), 656–673. <https://doi.org/10.1111/bjop.12292>

Scialfa, C. T., Borkenhagen, D., Lyon, J., Deschênes, M., Horswill, M., & Wetton, M. (2012). The effects of driving experience on responses to a static hazard perception test. *Accident Analysis & Prevention*, 45, 547–553. <https://doi.org/10.1016/j.aap.2011.09.005>

Song, J., Kosovicheva, A., & Wolfe, B. (2024). Road Hazard Stimuli: Annotated naturalistic road videos for studying hazard detection and scene perception. *Behavior Research Methods*, 56(4), 4188–4204. <https://doi.org/10.3758/s13428-023-02299-8>

Sulistyawati, K., & Chui, Y. P. (2009). Confidence Bias in Situation Awareness. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics* (Vol. 5639, pp. 317–325). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-02728-4_34

Sulistyawati, K., Wickens, C. D., & Chui, Y. P. (2011). Prediction in Situation Awareness: Confidence Bias and Underlying Cognitive Abilities. *The International Journal of Aviation Psychology*, 21(2), 153–174. <https://doi.org/10.1080/10508414.2011.556492>

Sun, L., Cheng, L., & Zhang, Q. (2021). The differences in hazard response time and driving styles of violation-involved and violation-free taxi drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 82, 178–186. <https://doi.org/10.1016/j.trf.2021.08.011>

Sun, L., Hu, W., Cheng, L., & Zhang, C. (2024). Effects of hazard type and confidence level on hazard perception in young male drivers: An ERP study. *NeuroReport*, 35(5), 299–305. <https://doi.org/10.1097/WNR.0000000000002007>

-
- Takahashi, R., Kobayashi, M., Sasaki, T., Yokokawa, Y., Momose, H., & Ohhashi, T. (2017). Driving simulation test for evaluating hazard perception: Elderly driver response characteristics. *Transportation Research Part F: Traffic Psychology and Behaviour*, 49, 257–270. <https://doi.org/10.1016/j.trf.2017.07.003>
- Wetton, M. A., Hill, A., & Horswill, M. S. (2011). The development and validation of a hazard perception test for use in driver licensing. *Accident Analysis & Prevention*, 43(5), 1759–1770. <https://doi.org/10.1016/j.aap.2011.04.007>
- Wetton, M. A., Horswill, M. S., Hatherly, C., Wood, J. M., Pachana, N. A., & Anstey, K. J. (2010). The development and validation of two complementary measures of drivers' hazard perception ability. *Accident Analysis & Prevention*, 42(4), 1232–1239. <https://doi.org/10.1016/j.aap.2010.01.017>
- White, M. J., Cunningham, L. C., & Titchener, K. (2011). Young drivers' optimism bias for accident risk and driving skill: Accountability and insight experience manipulations. *Accident Analysis & Prevention*, 43(4), 1309–1315. <https://doi.org/10.1016/j.aap.2011.01.013>
- Wohleber, R. W., & Matthews, G. (2016). Multiple facets of overconfidence: Implications for driving safety. *Transportation Research Part F: Traffic Psychology and Behaviour*, 43, 265–278. <https://doi.org/10.1016/j.trf.2016.09.011>
- Yan, F., Eilers, M., & Baumann, M. (2023). Investigating driver uncertainty about lane change decisions. *Transportation Research Part F: Traffic Psychology and Behaviour*, 95, 369–379. <https://doi.org/10.1016/j.trf.2023.05.001>
- Yan, F., Weber, L., & Luedtke, A. (2015). Classifying driver's uncertainty about the distance gap at lane changing for developing trustworthy assistance systems. *2015 IEEE Intelligent Vehicles Symposium (IV)*, 1276–1281.

<https://doi.org/10.1109/IVS.2015.7225858>

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1310–1321. <https://doi.org/10.1098/rstb.2011.0416>

Zeuwts, L. H. R. H., Vansteenkiste, P., Deconinck, F. J. A., Cardon, G., & Lenoir, M. (2017). Hazard perception in young cyclists and adult cyclists. *Accident Analysis & Prevention*, 105, 64–71. <https://doi.org/10.1016/j.aap.2016.04.034>

Zhang, J., Liao, Y., Wang, S., & Han, J. (2017). Study on Driving Decision-Making Mechanism of Autonomous Vehicle Based on an Optimized Support Vector Machine Regression. *Applied Sciences*, 8(1), 13. <https://doi.org/10.3390/app8010013>

Zhang, X., Chang, R., Wang, M., & Sui, X. (2024). The influence of driver's risk perception ability on driving decision-making: An ERP study. *Current Psychology*, 43(25), 21995–22005. <https://doi.org/10.1007/s12144-024-05884-y>

5

General Discussion

This thesis drew together empirical and theoretical knowledge about how individuals acquire perceptual information, how they evaluate the correctness of their perceptual performance, and assess risk in dynamic scenarios, from a human factors psychology standpoint. By integrating findings across three complementary studies, this work adopted a multi-component perspective on human performance in contexts requiring high visual engagement, considering cognitive processes (i.e., perception and adaptive decision-making), metacognitive monitoring (confidence judgments and risk estimation), and environmental factors as jointly relevant components.

5.1 Summary of the Main Findings

Chapter 2 provided a literature review on expertise-related differences in gaze behavior during visual search tasks, and the potential application of expert gaze behavior as an adjunct tool to support learning and task performance. It covered several theoretical accounts, skill transfer approaches, and eye-tracking research applied across different tasks and domains. Findings showed that expert gaze strategies are more selective, anticipatory, and task-relevant across domains. These patterns are not just the consequence of experience, but reflect tacit knowledge that can be externalized, studied, and partially transferred. Eye-tracking research highlighted how perceptual expertise supports selective attention and efficient cue integration. The review identified eight

empirical studies meeting strict inclusion criteria and found converging evidence that gaze-trained individuals outperformed those in traditional instruction groups. However, transfer effects were moderated by task complexity and prior knowledge, suggesting that gaze-based training is not universally effective but must be adapted to context-specific constraints.

In Chapter 3, the concepts of confidence judgments and their miscalibration in the context of driving hazard perception were discussed. Within this framework, confidence bias was conceptualized as the discrepancy between subjective confidence judgments regarding the correctness of one's task performance and their actual performance accuracy. Specific focus was given to understanding whether this bias affects both hazard perception and the subjective experience of risk. To this end, a new Hazard Perception Task (HPT) was developed, which integrated trial-by-trial confidence judgments and risk estimation of driving scenarios. Results revealed that overconfidence was associated with a lower probability of correct response in identifying hazards, without affecting reaction times. Hazard perception performance and risk estimation reflected interactions between hazard presence and lighting conditions. Confidence bias did not predict risk estimation, which appeared primarily driven by environmental factors. In addition, the association between Environmental Sensitivity (HSP scale) and subjective risk estimation in the presence of potential road hazards was explored in a preliminary manner, and showed a weak positive relationship.

In Chapter 4, the novel (driving) Hazard Perception Task developed in the previous study was implemented in a different national context, with data collected in the Netherlands from a predominantly Dutch sample (69.77%), and extended by integrating a (cycling) Decision-Making Task. This study examined how confidence judgments and risk estimation vary and interrelate across perception and decision-making processes, within different mobility modes (driving vs. cycling) and environmental conditions. Results showed cross-task consistency in confidence judgments and risk estimation, indicating that these evaluations generalize across different cognitive processes (perception and decision-making) and stimulus types (real-world driving scenes vs. simulated-urban cycling areas). In contrast, performance measures (response accuracy and braking behavior) did not correlate across tasks and

remained strongly dependent on task-specific situational environmental factors (e.g., nighttime conditions, pedestrian density). Also, Environmental sensitivity (HSP scale) showed a statistically significant but practically small effect on response accuracy in HPT. Although this result is consistent with previous work on sensitivity to contextual cues and threat awareness, its role in hazard detection warrants further investigation.

5.2 Theoretical Implications

The empirical findings of the thesis contribute to theoretical models of driving safety and situational awareness by highlighting the role of metacognitive components in hazard perception and decision-making. By anchoring a multi-component perspective in the human factors tradition and grounding it in recent cognitive science literature, the thesis argues for embedding self-monitoring processes such as confidence judgments and risk estimation into contemporary models of situational awareness and decision making.

Endsley's SA model (1995) remains foundational, but it describes the content of awareness, while metacognitive monitoring represents the assessment of that awareness, that is, for instance, the confidence with which one judges one's state of knowledge (Horrey et al., 2015). For example, McAnally and colleagues (2017) used a radar-like visual change-detection task to simulate an SA scenario and found that individuals often missed changes in the display (demonstrating limited SA) but nevertheless gave Judgment of Performance (JOP, that is perceived probability of a correct response) well calibrated to objective performance. In other words, even when situational awareness was imperfect, participants showed well-calibrated metacognitive insight into their performance. Embedding these monitoring processes into the SA model may help explain when and why perception alone fails to translate into safe or effective action. Achieving adaptive performance requires both accurate awareness and calibrated judgements.

5.3 Practical Implications

The findings have some practical implications.

First, the evidence presented in Chapter 2 on expert gaze behavior supports the development of training methods that integrate visual modeling into the teaching process. In high-risk settings that require intense visual engagement, exposing trainees to expert gaze patterns may help promote anticipatory and task-relevant attention strategies.

Second, as demonstrated in Chapters 3 and 4, assessment protocols in transportation, especially for driver licensing or professional certification, could benefit from integrating metacognitive indicators. Traditional hazard perception tests used in licensing procedures (Wetton et al., 2011) assess whether hazards are recognized, but they do not capture how sure individuals are about their judgments, nor how they evaluate risk. Including confidence and risk estimation alongside behavioral performance may offer a rich assessment of self-monitoring skills, as it may help identify individuals who systematically misjudge their abilities or underestimate risks (i.e., who are not well calibrated).

Third, the results support the development of adaptive support systems that detect misalignments between cognitive performance, confidence, and behavior. In situations consistent with “I looked, but I didn’t see it”, a system may register that visual information was sampled without leading to timely or appropriate action, indicating potential overconfidence, delayed awareness, or insufficient interpretation of available cues. By incorporating metacognitive indicators or gaze-based markers, such systems could move beyond behavior monitoring toward real-time decision support. Similar adaptive approaches may be relevant in domains such as aviation or healthcare, where misjudged certainty carries high costs and time for corrective action is limited.

5.4 Limitations and Future Directions

While the methodological limitations of each study have been discussed in their respective chapters, this thesis is also subject to some limitations that should be acknowledged. First, although the research adopts a multi-component perspective on human performance in high visual engagement scenarios (integrating cognitive, metacognitive, and individual components), it does this within a limited set of experimental paradigms. The use of controlled, computerized tasks in the laboratory has

ensured internal validity but has limited the exploration of higher-order interactions that typically emerge in more complex real-world, dynamic environments (Mackenzie & Harris, 2015). Future work should incorporate high-fidelity driving simulations or instrumented vehicles to assess the generalizability of these findings.

A second limitation concerns the reliance on trial-by-trial confidence ratings (CRs). While these provide valuable on-line monitoring insights, research shows they introduce cognitive load and may interfere with primary task performance (Petrusic & Baranski, 2003; Li et al., 2023). Specifically, CRs can lead to longer response times, enhanced decision accuracy, and compete for attentional resources. Future studies should include implicit markers of confidence, such as decision latency, or physiological measures like pupil dilation (Fleming and Dolan, 2012; Lempert et al., 2015).

Third, the samples of participants in the experiments were predominantly composed of young adults. This limits the generalizability of the results to other populations, such as older drivers, professional operators, or individuals from different cultural backgrounds. For example, previous research suggests that aging may affect both confidence calibration and perceptual performance (Klever et al., 2022), a limit that should be the subject of future research.

Finally, this thesis only preliminarily addressed the role of individual traits such as Environmental Sensitivity. Future studies should investigate more how this factor interacts with contextual demands to shape perceptual performance and threat awareness (Rubaltelli et al., 2018).

References

- Endsley, M. R. (1995). Measurement of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 65–84. <https://doi.org/10.1518/001872095779049499>
- Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1338–1349. <https://doi.org/10.1098/rstb.2011.0417>
- Horrey, W. J., Lesch, M. F., Mitsopoulos-Rubens, E., & Lee, J. D. (2015). Calibration of skill and judgment in driving: Development of a conceptual framework and the implications for road safety. *Accident Analysis & Prevention*, 76, 25–33. <https://doi.org/10.1016/j.aap.2014.12.017>
- Klever, L., Mamassian, P., & Billino, J. (2022). Age-related differences in visual confidence are driven by individual differences in cognitive control capacities. *Scientific Reports*, 12(1), 6016. <https://doi.org/10.1038/s41598-022-09939-7>
- Lempert, K. M., Chen, Y. L., & Fleming, S. M. (2015). Relating Pupil Dilation and Metacognitive Confidence during Auditory Decision-Making. *PLOS ONE*, 10(5), e0126588. <https://doi.org/10.1371/journal.pone.0126588>
- Li, B., Hu, X., Shanks, D. R., Su, N., Zhao, W., Meng, L., Lei, W., Luo, L., & Yang, C. (2024). Confidence ratings increase response thresholds in decision making. *Psychonomic Bulletin & Review*, 31(3), 1093–1102. <https://doi.org/10.3758/s13423-023-02380-5>
- Mackenzie, A. K., & Harris, J. M. (2015). Eye movements and hazard perception in active and passive driving. *Visual Cognition*, 23(6), 736–757. <https://doi.org/10.1080/13506285.2015.1079583>

- McAnally, K. I., Morris, A. P., & Best, C. (2017). Metacognitive monitoring and control in visual change detection: Implications for situation awareness and cognitive control. *PLOS ONE*, 12(9), e0176032. <https://doi.org/10.1371/journal.pone.0176032>
- Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgments. *Psychonomic Bulletin & Review*, 10(1), 177–183. <https://doi.org/10.3758/BF03196482>
- Rubaltelli, E., Scrimin, S., Moscardino, U., Priolo, G., & Buodo, G. (2018). Media exposure to terrorism and people's risk perception: The role of environmental sensitivity and psychophysiological response to stress. *British Journal of Psychology*, 109(4), 656–673. <https://doi.org/10.1111/bjop.12292>
- Wetton, M. A., Hill, A., & Horswill, M. S. (2011). The development and validation of a hazard perception test for use in driver licensing. *Accident Analysis & Prevention*, 43(5), 1759–1770. <https://doi.org/10.1016/j.aap.2011.04.007>

Supplementary Materials Section A - Experiments Materials

Table A1. Common Eye-Tracking Measures and Definitions.

Category	Measure	Definition
Position Measures	Position of a fixation	Addresses where a participant looks, i.e., the center of the field of vision. It is given by (x,y)-coordinates in a two-dimensional space, or (x,y,z)-coordinates in a three-dimensional space.
Position Measures	Fixation duration (also known as fixation time)	Measures the time the gaze of a participant stays in a position.
Position Measures	Dwell time	Measures the sum of all fixation durations during a dwell (i.e., a visit) in an AOI, from entry to exit
Position Measures	Pupil diameter (also known as pupil dilation or pupil size)	Quantifies the diameter of the pupil. Its values are normally given in the form of pixels of the eye camera or in millimeters after calibration
Numerosity measures	Number of fixations (also known as fixation density)	Quantifies how often an AOI is fixated.

Category	Measure	Definition
Numerosity measures	Fixation rate (also known as fixation frequency)	Measures the number of fixations per second or per minute.
Numerosity measures	Blink rate	Measures the number of blinks per second or per minute.
Movement measures	Saccade amplitude (also known as saccade length or saccade distance)	Measures the distance of eye movement from one fixation to a consecutive fixation.
Movement measures	Saccade duration (also known as transition time)	Measures the time the saccade takes to move between two fixations.
Movement measures	Saccade velocity	Refers to the first derivation of position data with respect to time.
Aggregate measures	Attention map	It is normally visualized using a heatmap and describes the spatial distribution of eye movement data.
Aggregate measures	Scanpath (also known as scan pattern, search pattern, gaze sequence)	Describes how the eye moves physically through space within a certain timespan.

Category	Measure	Definition
Aggregate measures	Saccade amplitude fixation duration ratio (SF ratio)	Quantifies the proportional measure of saccade amplitude to fixation duration.

Table A2. Categorical and Temporal Annotation of Video-based Hazard Perception Task Stimuli.

The Hazard Perception Task consisted of 80 videos (selected from Song et al., 2024). Table A2 reports the categorical and temporal annotation of the stimuli, with a legend provided below.

Legend for Table A2:

Annotation Name	Description
Hazard / Safe	Whether a hazard is present in the video
day / night	Time of day the video is taken
MovieFile_name	Name of the file
Duration (s)	Duration of the video
Hazard onset (s)	Time in seconds of hazard onset from the start of video, i.e. the initial indication that the situation is hazardous, not the first moment the eventual hazardous object becomes visible.
Type of hazard	Category of hazard: vehicle, pedestrian, cyclist, animal

Annotation Name	Description
Hazard side	The location of the hazard relative to video midline
Hazard Frame	Screenshot representing the hazard
Brief description	Brief description of how the hazard develops
Driver response time	Time in seconds of first driver response from start of video
Onset_response	Time in seconds between hazard onset and video driver response

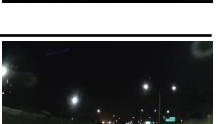
Table A2.

Trial	Hazard / Safe	Day / Night	MovieFile_name	Duration (s)	Hazard onset (s)	Type of Hazard	Hazard side	Brief description	Driver response time	Onset_res ponse	Video Frame
1	Hazard	day	noaudio_c_002_critical_pedestrian	8.16	4.7	vehicle	left	Car's door opens along the adjacent lane	5.23	0.53	
2	Hazard	day	noaudio_c_008_critical_driver_hits_sign	8.0	3.5	vehicle	right	Dynamic obstacles caused by a vehicle going off the road	7.9	4.4	
3	Hazard	day	noaudio_c_008_critical_traffic_cone	8.14	2.45	obstacle	right	Dynamic obstacles (a traffic cone) crossing the roadway	3.42	0.95	
4	Hazard	day	noaudio_c_010_critical_unsignaled_turn	8.0	5.05	vehicle	left	Unsignaled turn by oncoming vehicle in lane	5.4	0.35	
5	Hazard	day	noaudio_c_012_critical_cut_across	8.04	3.93	vehicle	right	Vehicle pop out that performs a turnaround	5.7	1.77	

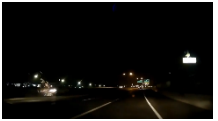
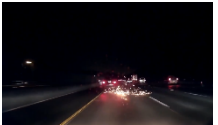
6	Hazard	day	noaudio_c_013_c ritical_u_turn	8.03	3.7	vehicle	right	Oncoming vehicle in lane that performs a turnaround	4.33	0.63	
7	Hazard	day	noaudio_c_016_c ritical_loose_tires	8.0	0.91	obstacles	left	Dynamic obstacles (two tires) crossing the roadway in the car's direction in snowy conditions	2.46	1.56	
8	Hazard	day	noaudio_c_018_c ritical_swerve	8.03	3.75	vehicle	left	Critical swerve by a vehicle that is not visible because it is immediately behind another.	5.45	1.7	
9	Hazard	day	noaudio_c_019_c ritical_swerve	8.08	2.52	vehicle	right	Critical entry of a vehicle into our lane in rainy conditions	3.3	0.78	
10	Hazard	day	noaudio_c_020_c ritical_deer	8.0	5.55	animal	left	A deer suddenly crosses the roadway in rain conditions	5.76	0.22	
11	Hazard	day	noaudio_c_022_c ritical_truck_crashes	8.06	3.14	vehicle	right	Critical truck crash	6.7	3.56	

12	Hazard	day	noaudio_c_023_c ritical_truck_cras hes	8.03	2.83	vehicle	right	Critical incident in which a truck overturns	4.4	1.57	
13	Hazard	day	noaudio_c_024_c ritical_stopped_c ar_in_road	8.0	2.03	vehicle	left	A stopped car occupies the lane of traffic	2.3	0.27	
14	Hazard	day	noaudio_c_026_c ritical_run_light	7.97	5.33	vehicle	right	A vehicle suddenly arriving from the right cutting us off	5.5	0.17	
15	Hazard	day	noaudio_c_029_c ritical_swerve	8.04	4.25	vehicle	right	Critical swerve of a vehicle for lane change	6.0	1.75	
16	Hazard	day	noaudio_c_035_c ritical_crash_in_l ane	8.0	3.3	vehicle	left	Critical lane crash, involved three vehicles	4.97	1.66	
17	Hazard	day	noaudio_c_045_c ritical_tire	8.07	1.4	obstacle	left	Dynamic obstacles (a tire) crossing the roadway in the car's direction	3.7	2.3	

SUPPLEMENTARY MATERIALS

18	Hazard	day	noaudio_c_046_critical_pedestrian	8.0	4.7	pedestrian	right	Pedestrian crossing right of way	6.5	1.8	
19	Hazard	day	noaudio_c_critical_001	8.0	4.0	vehicle	left	Critical lane crash, involved two vehicles	4.73	0.73	
20	Hazard	day	noaudio_c_critical_055	8.03	6.0	animal	left	An animal suddenly crosses the roadway in snowy conditions	7.25	1.25	
21	Hazard	night	noaudio_c_critical_070	8.0	5.0	vehicle	right	Critical entry of a vehicle into our lane	6.0	1.0	
22	Hazard	night	noaudio_c_critical_072	8.06	4.57	animal	right	An animal suddenly crosses the roadway in snowy conditions	5.15	0.58	
23	Hazard	night	noaudio_c_critical_162	8.03	4.7	vehicle	right	Crash collision of a vehicle against the guardrails	5.65	0.95	

SUPPLEMENTARY MATERIALS

24	Hazard	night	noaudio_c_critical_163	8.04	1.37	vehicle	left	Crash crash of a vehicle against the security wall	2.84	1.48	
25	Hazard	night	noaudio_c_critical_186	8.0	4.0	obstacles	/	Dynamic obstacles crossing the roadway in the car's direction	4.6	0.6	
26	Hazard	night	noaudio_c_critical_200	8.0	3.0	vehicle	left	Oncoming vehicle in lane	3.5	0.5	
27	Hazard	night	noaudio_c_critical_209	8.06	2.6	vehicle	right	Critical swerve of a vehicle for lane change	3.17	0.57	
28	Hazard	night	noaudio_c_critical_232	8.03	6.0	vehicle	left	Vehicle breaks into the lane of traffic, cutting us off	6.85	0.75	
29	Hazard	night	noaudio_c_critical_266	8.04	1.0	pedestrian	right	Pedestrian crossing right of way	3.3	2.3	

30	Hazard	night	noaudio_c_critical_276	8.04	5.8	animal	right	An animal suddenly crosses the roadway in snowy conditions	6.55	0.75	
31	Hazard	night	noaudio_c_critical_307	8.3	4.5	vehicle	left	Vehicle breaks into the lane of traffic, cutting us off	5.98	1.48	
32	Hazard	night	noaudio_c_critical_315	8.0	4.61	vehicle	right	Critical swerve of a vehicle for lane change	5.03	0.42	
33	Hazard	night	noaudio_c_critical_319	8.0	3.22	obstacle	left	Dynamic obstacles (a tire) crossing the roadway in the car's direction	3.72	0.5	
34	Hazard	night	noaudio_c_critical_352	8.0	2.35	vehicle	right	Critical swerve of a vehicle for lane change	3.08	0.73	
35	Hazard	night	noaudio_c_critical_359	8.0	2.5	animal	right	An animal stopped on the roadway	4.03	1.53	

SUPPLEMENTARY MATERIALS

36	Hazard	night	noaudio_c_critical_369	8.0	3.56	animal	right	An animal suddenly crosses the roadway	4.25	0.69	
37	Hazard	night	noaudio_c_critical_395	7.96	1.0	vehicle	right	Loss of control of a vehicle in the adjacent lane in snowy conditions	2.28	1.28	
38	Hazard	night	noaudio_c_critical_405	7.96	1.52	vehicle	right	Vehicle breaks into the lane of traffic, cutting us off in snowy conditions	2.34	0.82	
39	Hazard	night	noaudio_c_critical_417	8.0	3.6	vehicle	right	Crash crash of a vehicle against the security wall	4.2	0.61	
40	Hazard	night	noaudio_c_critical_435	8.04	5.7	pedestrian	left	Pedestrian crossing right of way	6.55	0.85	
41	Safe	day	noaudio_c_005_no event	10.07							
42	Safe	day	noaudio_c_006_no event 2	8.0							
43	Safe	day	noaudio_c_012_no event 1	8.01							

44	Safe	day	noaudio_c_013_n o event 1	8.03
45	Safe	day	noaudio_c_016_n o event 1	8.01
46	Safe	day	noaudio_c_018_n o event 1	8.0
47	Safe	day	noaudio_c_019_n o event 1	8.08
48	Safe	day	noaudio_c_024_n o event 1	8.03
49	Safe	day	noaudio_c_033_n o event 1	6.0
50	Safe	day	noaudio_c_036_n o event 2	8.03
51	Safe	day	noaudio_c_039_n o event 2	10.27
52	Safe	day	noaudio_c_041_n o event 1	8.0
53	Safe	day	noaudio_c_041_n o event 3	8.07
54	Safe	day	noaudio_c_044_n o event 1	8.08
55	Safe	day	noaudio_c_046_n o event 1	8.08
56	Safe	day	noaudio_c_047_n o event 1	8.0
57	Safe	day	noaudio_c_049_n o event 2	8.03
58	Safe	day	noaudio_c_049_n o event 3	8.0

59	Safe	day	noaudio_c_050_n o_event_3	8.0
60	Safe	day	noaudio_c_054_n o_event_2	8.04
61	Safe	night	noaudio_c_no_ev ent_011	8.1
62	Safe	night	noaudio_c_no_ev ent_070	8.14
63	Safe	night	noaudio_c_no_ev ent_072	8.0
64	Safe	night	noaudio_c_no_ev ent_084	8.0
65	Safe	night	noaudio_c_no_ev ent_090	8.02
66	Safe	night	noaudio_c_no_ev ent_098	8.07
67	Safe	night	noaudio_c_no_ev ent_125	8.03
68	Safe	night	noaudio_c_no_ev ent_125b	8.0
69	Safe	night	noaudio_c_no_ev ent_163	8.01
70	Safe	night	noaudio_c_no_ev ent_186	8.01
71	Safe	night	noaudio_c_no_ev ent_200	8.03
72	Safe	night	noaudio_c_no_ev ent_209	8.1
73	Safe	night	noaudio_c_no_ev ent_291	2.51

74	Safe	night	noaudio_c_no_ev ent_291b	4.93
75	Safe	night	noaudio_c_no_ev ent_317	4.23
76	Safe	night	noaudio_c_no_ev ent_323	1.75
77	Safe	night	noaudio_c_no_ev ent_334b	10.6
78	Safe	night	noaudio_c_no_ev ent_359	5.8
79	Safe	night	noaudio_c_no_ev ent_362	5.0
80	Safe	night	noaudio_c_no_ev ent_376	7.11

Reference:

Song, J., Kosovicheva, A. & Wolfe, B. Road Hazard Stimuli: Annotated naturalistic road videos for studying hazard detection and scene perception. *Behav Res* **56**, 4188–4204 (2024). <https://doi.org/10.3758/s13428-023-02299->

Supplementary Material A3. Demographic and Mobility-related Experience Questionnaire

Participants completed self-reported questionnaires assessing demographic variables (age, gender identity, education) and mobility-related experience, including years since obtaining a driver's license, driving frequency and mileage in the past 12 months, and cycling history (i.e., age of acquisition and current cycling frequency per month).

The full questionnaire is provided below.

Q1. What is your age? (insert number) _____

Q2. What gender do you self-identify as?

- ☐ Man
- ☐ Woman
- ☐ Non-binary
- ☐ Prefer to self-describe as: _____
- ☐ Prefer not to answer

Q3. What is the highest degree that you have already completed / finished? Please select one of the following alternatives:

- ☐ High school diploma
- ☐ Bachelor's degree
- ☐ Master's degree
- ☐ PhD or post-graduate degree
- ☐ Other, namely: _____

Q4. How many years do you have a driver's license? _____

Q5. On average, how often did you drive during the past 12 months? Please select the closest answer:

- ☐ Once a week or less
- ☐ More than twice a week

Q6. How many kilometers have you driven in the past 12 months roughly? Please select the closest answer:

- ☐ Less than 5,000 km per year

- ☐ More than 10,000 km per year

Q7. Do you drive any other vehicles? Please tick all that apply

- ☐ Bicycle
- ☐ Scooter/Motorcycle
- ☐ No other vehicle
- ☐ Other (please specify) _____

Q8. If you ride a bicycle, at what age did you learn to ride one?

- ☐ 0-9 years old
- ☐ 10-18 years old
- ☐ Over 18 years old

Q9. If you ride a bicycle, how often do you cycle in a typical month?

- ☐ Almost every day

- ☐ 3-4 times a week
- ☐ Once a week
- ☐ Once a month
- ☐ Rarely or never

Q10. Where did you grow up or spend most of your time?

- ☐ Netherlands
- ☐ Another country (please specify) _____
- ☐ Prefer not to answer

Supplementary Material A4. General Confidence Questionnaire

At the end of the experiment, participants completed a final 3-item questionnaire assessing general confidence. The items assessed confidence in task performance, confidence in traffic participation, and general self-confidence, each rated on a 7-point scale (1 = “Not at all,” 7 = “Extremely”).

The full general confidence questionnaire is provided below.

Q1. Across the two tasks that you did today (Hazard Perception Task, Cycling Task): How confident are you that you performed well across these two tasks?

☒ 1	☒ 2	☒ 3	☒ 4	☒ 5	☒ 6	☒ 7
(not at all)						(extremely confident)

Q2. In general, beyond the experiment you did today: How confident are you in your overall ability to participate in traffic in general?

☐	☐	☐	☐	☐	☐	☐
1	2	3	4	5	6	7
(not at all)						(extremely confident)

Q3. In general, beyond the experiment you did today and beyond just traffic settings: How would you rate your general confidence?

☐	☐	☐	☐	☐	☐	☐
1	2	3	4	5	6	7
(not at all)						(extremely confident)

Supplementary Material A5. Confidence Bias Formulas

These formulas define Hazard Perception Accuracy, Confidence Judgments, and Confidence Bias, and were used to derive individual-level measures and classify participants as overconfident or not overconfident.

$$1. \text{ Hazard Perception Accuracy \%} = \frac{\text{Sum of Score}}{\text{Total completed trials}} * 100$$

$$2. \text{ Confidence Judgments \%} = (\text{Mean Confidence between trials} - 1) * \frac{100}{6}$$

$$3. \text{ Confidence Bias} = \text{Confidence Judgments \%} - \text{Hazard Perception Accuracy \%}$$

$$4. \text{ Groups} = \text{IF Confidence Bias} > 0 \text{ THEN "Overconfident", ELSE "Not Overconfident"}.$$

Supplementary Materials Section B - Full Model Estimates

Table B1. GLMM fixed-effects estimates for Response Accuracy

	Estimate	SE	t-score	p-value	
(Intercept)	2.69	0.07	36.28	<.001	***
Confidence bias	-0.04	0.01	-4.42	<.001	***
Hazard	0.21	0.06	3.43	<.001	***
Nighttime	-0.29	0.06	-4.82	<.001	***
Confidence bias x Hazard	0.005	0.009	0.55	.581	
Confidence bias x Daytime	-0.003	0.009	-0.35	.729	
Hazard x Daytime	-0.41	0.06	-6.79	<.001	***

Significance: *: $p < .05$; **: $p < .01$; ***: $p < .001$.

Table B2. LMM fixed-effects estimates for Reaction Times (RTs)

	Estimate	SE	df	t-score	p-value	
(Intercept)	826.12	46.00	62.96	17.96	<.001	***
Confidence bias	-4.46	5.88	63.01	-0.76	.451	
Nighttime	-98.83	17.41	2387.29	-5.68	<.001	***

*Significance: *: $p < .05$; **: $p < .01$; ***: $p < .001$.*

Table B3. LMM fixed-effects estimates for Subjective Risk Estimation

	Estimate	SE	df	t-score	p-value	
(Intercept)	3.79	0.08	63.00	50.52	<.001	***
Confidence bias	-0.004	0.01	63.02	-0.41	.683	
Hazard	1.77	0.02	5103.03	93.87	<.001	***
Nighttime	0.17	0.02	5103.03	8.78	<.001	***
Confidence bias x Hazard	0.01	0.002	5103.03	5.50	<.001	***
Confidence bias x Daytime	-0.0004	0.002	5103.03	-0.18	.860	
Hazard x Daytime	0.15	0.02	5103.03	8.25	<.001	***

Significance: *: $p < .05$; **: $p < .01$; ***: $p < .001$.

Supplementary Material B4. Results Balloon Analogue Risk Task (BART)

The average percentage of balloons exploded per participant was 29% (SD = 16%).

Table B5. Results Driving Hazard Perception Task: Full model estimates Accuracy score

	Estimate	SE	t-score	p-value	
(Intercept)	10.59	2.10	5.05	<.001	***
Night	-4.54	1.97	-2.30	.021	*
Hazard	-10.97	2.23	-4.92	<.001	***
HSP	-1.32	0.41	-3.26	.001	**
HPQ1	-0.03	0.18	-0.18	.856	
Previous conf	-0.12	0.07	-1.60	.110	
General conf2	0.02	0.13	0.14	.892	
General conf3	-0.07	0.14	-0.48	.629	
Night: Hazard	5.51	2.61	2.11	.035	*
Night: HSP	0.68	0.41	1.67	.095	
Hazard: HSP	2.28	0.484	4.71	<.001	***
Night: Hazard: HSP	-0.89	0.573	-1.56	.119	

Table B6. Results Driving Hazard Perception Task: Full model estimates Response times in hazardous trials

	Estimate	SE	t-score	p-value	
(Intercept)	0.77	0.19	4.13	<0.001	***
Night	-0.091	0.030	-3.01	0.003	**
HSP	-0.041	0.031	-1.32	0.187	
HPQ1	-0.012	0.026	-0.48	0.631	
Previous conf	0.011	0.012	0.91	0.363	
General conf2	0.0020	0.019	0.10	0.917	
General conf3	0.025	0.019	1.28	0.202	

Table B7. Results Driving Hazard Perception Task: Full model estimates Subjective Risk Estimation (All trials)

	Estimate	SE	t-score	p-value	
(Intercept)	3.68	0.76	4.79	<.001	***
Accuracy score	-1.51	0.17	-9.10	<.001	***
Night	0.37	0.17	2.18	.030	*
Hazard	0.20	0.18	1.13	.260	
HSP	0.10	0.13	0.79	.432	
HPQ	-0.22	0.11	-1.98	.048	*
Previous conf	0.01	0.02	0.41	.683	
General conf2	-0.05	0.08	-0.63	.528	
General conf3	0.062	0.09	0.72	.472	
Night: accuracy	-0.095	0.18	-0.54	.590	
Hazard: accuracy	3.57	0.18	19.29	<.001	***

Table B8. Results Driving Hazard Perception Task: Full model estimates Subjective Risk Estimation (Hazard trials)

	Estimate	SE	t-score	p-value	
(Intercept)	5.53	0.95	5.85	<0.001	***
Night	0.20	0.15	1.40	0.161	
Response Time	-0.005	0.11	-0.042	0.967	
HSP	0.09	0.16	0.57	0.567	
HPQ	-0.17	0.14	-1.23	0.220	
Previous conf	0.05	0.033	1.39	0.166	
General conf2	-0.05	0.10	-0.50	0.620	
General conf3	0.09	0.11	0.86	0.389	
Night: Response Time	-0.38	0.18	-2.08	0.0377	*

Table B9. Results Driving Hazard Perception Task: Full model estimates Local confidence judgements (all trials)

	Estimate	SE	t-score	p-value	
(Intercept)	2.72	0.83	3.29	0.001	**
Subjective Risk Estimation	0.48	0.10	4.77	<.001	***
Accuracy score	3.37	0.36	9.30	<.001	***
Night	1.45	0.39	3.73	<.001	***
Hazard	0.89	0.40	2.20	.038	*
HSP	-0.07	0.13	-.56	.575	
HPQ	0.16	0.11	1.40	.161	
Previous conf	0.05	0.02	2.73	.006	**
General conf2	0.05	0.08	0.59	.553	
General conf3	-0.01	0.09	-0.06	.949	
Subjective Risk Estimation: Night	-0.40	0.11	-3.82	<.001	***
Subjective Risk Estimation: Hazard	-0.42	0.11	-3.83	<.001	***
Subject Risk Estimation: Accuracy score	-0.87	0.10	-8.44	<.001	***
Hazard: Accuracy score	-2.38	0.40	-4.30	<.001	***

SUPPLEMENTARY MATERIALS

Risk: night: accuracy score	0.45	0.44	-5.38	<.001	***
Risk: hazard: accuracy score	1.00	0.11	4.16	<.001	***

Table B10. Results Driving Hazard Perception Task: Full model estimates Local confidence judgements (hazard trials)

	Estimate	SE	t-score	p-value	
(Intercept)	-0.29	0.88	-0.32	0.74	***
Subjective Risk Estimation	0.87	0.11	7.95	<.001	***
Night	-0.14	0.24	-0.59	.554	
Response Time	0.29	0.29	1.02	.309	
HSP	0.09	0.09	0.92	.359	
HPQ1	1.07	0.17	6.09	<.001	***
Previous conf	0.05	0.02	2.26	.024	*
General conf2	0.00	0.06	0.01	.991	
General conf3	0.03	0.06	0.55	.584	
Risk estimation: HPQ	-0.15	0.02	-5.94	<.001	***
Night: HPQ	0.02	0.06	0.40	.691	
Response time: HPQ	-0.05	0.07	-0.75	.453	

Table B11. Results Cycling Decision-Making Task: Full model estimates Braking behaviour

	Estimate	SE	t-score	p-value	
(Intercept)	1.37	3.26	0.42	.669	
Low pedestrian density	-0.49	0.14	-3.63	<.001	***
TTA	-0.35	0.068	-5.12	<.001	***
HSP	0.27	0.55	0.49	.627	
HPQ	0.87	0.49	1.78	.076	
Previous conf	0.098	0.052	1.89	.059	
General conf2	-0.24	0.35	-0.69	.491	
General conf3	-0.16	0.37	-0.43	.669	

Table B12. Results Cycling Decision-Making Task: Full model estimates Response time

	Estimate	SE	t-score	p-value	
(Intercept)	-0.76	1.12	-0.68	.500	
Low pedestrian density	-0.57	0.33	-1.70	.089	
TTA	0.20	0.048	4.11	<.001	***
HSP	0.23	0.19	1.20	.232	
HPQ	0.30	0.17	1.80	.072	
Previous conf	-0.0092	0.011	-0.83	.408	
General conf2	-0.066	0.12	-0.54	.590	
General conf3	0.039	0.13	0.24	.812	
Brake	-1.03	0.27	-3.76	<.001	***
Low pedestrian density: TTA	0.12	0.063	1.93	.054	
Low pedestrian density: brake	-0.55	0.36	-1.54	.124	
TTA: brake	0.28	0.052	5.38	<.001	***
Low pedestrian density: TTA: brake	0.14	0.069	2.10	.036	

Table B13. Results Cycling Decision-Making Task: Full model estimates Subjective Risk Estimation

	Estimate	SE	t-score	p-value	
(Intercept)	3.51	1.95	1.80	0.0717	
Low pedestrian density	-0.92	0.19	0.88	<.001	***
TTA	-0.094	0.10	-4.92	.358	
Response time	1.072	0.30	-0.92	.003	**
HSP	-0.0083	0.32	3.63	.980	
HPQ	-0.15	0.28	-0.026	.608	
Previous conf	0.015	0.17	-0.51	.381	
General conf2	-0.00017	0.21	-0.00084	1.00	
General conf3	-0.16	0.22	-0.73	.465	
Brake	1.72	0.67	2.57	.010	*
Low pedestrian density: Response time	0.289	0.081	3.55	<.001	***
TTA: response time	-0.12	0.051	-2.39	.017	*
Low pedestrian density: brake	0.21	0.22	1.01	.312	
TTA: brake	-0.043	0.12	-0.35	.729	

SUPPLEMENTARY MATERIALS

Response time: brake	-0.55	0.32	-1.70	.089
Pedestrian density low: Response time: brake	-0.081	0.090	-0.91	.363
TTA: response time: brake	0.068	0.56	1.22	.224

Table B14. Results Cycling Decision-Making Task: Full model estimates Local confidence judgements

	Estimate	SE	t-score	p-value	
(Intercept)	6.18	1.04	5.92	<.001	***
Low pedestrian density	0.80	0.19	4.32	<.001	***
TTA	-0.14	0.10	-1.42	.153	
Response time	-0.66	0.30	-2.22	.037	*
Subjective Risk Estimation	-0.58	0.075	-7.75	<.001	***
HSP	0.040	0.15	0.26	.792	
HPQ	0.18	0.13	1.34	.181	
Previous confidence	0.059	0.017	3.52	<.001	***
General confidence driving	-0.017	0.097	-0.18	.861	
General confidence overall	0.072	0.10	0.72	.472	
Brake	-2.46	0.69	-3.55	<.001	***
Low pedestrian density: Response time	-0.51	0.081	-6.35	<.001	***
TTA: Response time	0.16	0.050	3.15	.002	**
Response time: Risk estimate	-0.010	0.27	-0.39	.695	

SUPPLEMENTARY MATERIALS

Density low: Brake	-1.20	0.22	-5.58	<.001	***
Brake: TTA	0.19	0.12	1.62	.106	
Brake: Response time	0.92	0.32	2.83	.005	*
Brake: Risk estimate	0.73	0.083	8.85	<.001	***
Density low: Response time: brake	0.67	0.088	7.59	<.001	***
TTA:Response time: brake	-0.21	0.055	-3.84	<.001	***
Response time: risk estimate: brake	0.015	0.029	0.51	.613	
