Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Behavioral evaluation of two roadside warning formats in immersive virtual reality for landslide-risk driving

Arjun Mehra
Indian Institute of Technology Mandi
Kulbhushan Chand
Indian Institute of Technology Mandi
Tad Gonsalves
Sophia University
Joachim Meyer
Tel Aviv University
Akash K. Rao
Manipal Academy of Higher Education
Kala Venkata Uday
Indian Institute of Technology Mandi
Varun Dutt
varun@iitmandi.ac.in

Indian Institute of Technology Mandi

Article

Keywords: Virtual reality, Warning design, Driving behaviour, Human factors, User experience, Extended reality

Posted Date: September 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10825519/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. All authors declare no competing interests. Varun Dutt is a Guest Editor of the Scientific Reports Collection to which this manuscript is intended to be

submitted and will take no part in the editorial handling or peer-review decision for this manuscript; editorial management should be assigned independently in accordance with journal policy.

Abstract

Roadside warning systems communicate hazard information so drivers can initiate safety precautions, yet evidence linking warning formats to behavioural responses on landslide-prone roads remains limited. We compared an active visual warning with a passive pictographic landslide sign in an immersive virtual-driving simulation. Thirty-five participants were randomly allocated to an active-cue group (n = 17) or passive-cue group (n = 18) and completed two trials on rainy mountain roads while vehicle telemetry was recorded continuously. The primary analysis compared participant-level mean speed during the 10 s after the combined cue–hazard trigger in Attempt 1. The active-cue group drove more slowly than the passive-cue group (21.6 vs 32.6 km/h; mean difference, − 11.0 km/h; 95% CI, − 21.5 to − 0.5; Cohen's d = 0.91; p = .041). The same pattern appeared over 20 s (17.4 vs 29.8 km/h; p = .007). Warning format was also associated with changes in trigger speed across attempts (p = .023). These exploratory findings indicate different speed-regulation responses to the two intact warning formats, without isolating the effect of any individual visual component.

Introduction

Landslide hazards occur rapidly, threatening traffic safety on mountainous routes [1]. During precipitation or reduced visibility, drivers on landslide-prone roads face tight temporal constraints on perception and interpretation of imminent dangers [1], [2]. An early-warning system, as part of disaster-risk management, offers protection potential, conditional upon drivers being informed of the danger and translating such knowledge into protective behaviour [2–4]. From a human-centered user-experience perspective, therefore, a roadside warning serves as a critical safety interface for transferring technical information into useful behaviour by drivers. Information technology plays a vital role in providing information for decision-making; however, behavioural adaptations determine whether drivers can benefit from the available information or not [4].

Experimental tests with real-world imminent landslide scenarios face significant challenges due to unpredictability, ethical issues, and uncontrollable conditions in hazardous events. Immersive virtual reality (VR) allows creating experimental hazard scenarios under controlled settings while measuring participants' continuous response behaviour [5–8]. Recent systematic reviews of the application of immersive virtual/augmented reality to road safety suggest that these technologies emerge increasingly useful in driver training, assistance, and behaviour evaluation. However, there remain gaps in evaluating interfaces, populations, and transferring knowledge to actual driving behaviour [20]. Studies on road hazard communication and simulated driving behaviour have explored the impact of static and dynamic warnings on driver behaviour and selective attention using behavioural and physiological measurements such as reaction time, braking, gaze, and handover manoeuvvrves [9–12], [21]. Within landslide hazard communication literature, immersive virtual reality has been employed to simulate debris flow hazards and assess risk perception, warning reliability, collision behaviour, velocity, and route selection [7], [13], [14]. Nevertheless, relatively limited literature examines and compares distinct warning formats for imminent landslides using vehicle telemetry within the immersive landslide-driving simulation setting.

Any roadside active warning requires an upstream hazard-detection protocol capable of detecting and classifying an event that needs warning transmission. Operational landslide warning protocols rely upon threshold criteria based upon rainfall intensity-duration, soil moisture and groundwater piezometry, displacement rates, acoustic emission, among others. Warning lead-times vary according to failure mechanisms; thus, the time horizon for effective warning behaviour varies [2], [3]. For simplicity, in the current experiment, the warning protocol operates under full efficiency without producing a single missed detection nor false alarm. It is a modelling assumption rather than empirical validation because drivers tend to discount compliance when warning reliability decreases. Separate experiments have tested reliability manipulations in previous landslide hazard VR research by our lab [15]. Thus, this study assesses drivers' responsiveness to an assumed perfectly reliable warning downstream of a hypothetical hazard detection protocol; this study neither validates the operational prediction of imminent landslides, sensor accuracy nor warning lead time.

A warning's success cannot be attributed solely to information content. Conspicuousness arises out of several variables such as luminance, chromaticity, dimensions, eccentricity, and onset dynamics, whereby an abrupt visual stimulus attracts selective attention. Simulated driving behaviour experiments have shown that animated or dynamic cues increase detectability and improve timing judgements. However, braking performance depends upon visual stimuli and driving conditions [10], [12], [21]. Previous findings indicate that conventional triangular warning symbols might be easily recognised due to high familiarity for drivers, whereas state-changing illumination is less common. Therefore, if one type of warning format outperforms another in the current experiment, it implies the combination of temporal informativeness, conspicuity, novelty, and semantics [10], [12], [16], [21].

Building on this context, we aim to answer the specific question: what would happen if we randomly compare warning interfaces within an immersive landslide-hazard driving task? What is the nature of the difference in driver speed regulation between static vs dynamic warnings embedded within a landslide driving simulation? The novelty of this study lies in the random allocation of participant groups exposed to warning cues, the immersion of a two-route task environment that maintains participants' elected exposure level, and vehicle telemetry as a continuous measure of driver speed regulation along two driving tracks. Further novelty resides in making an explicit distinction between whole warning format comparisons and separate examination of individual visual components constituting such warning format [7], [9–14], [20], [21]. Our study aims to quantify the difference between driver speed regulation for two randomly allocated warning interface formats within the same landslide hazard driving task, examining consistency in the directional pattern over multiple repetitions and describing selected route choices, task completion duration, and subjective questionnaires.

Our hypothesis derives from the Triad of Risk-Related Behaviour framework, which conceptualises hazard-related decisions as a triadic relationship between exposure to a risk source, utilisation of protective features, and behavioural response to warning cues [17]. Exposure represents voluntary engagement in risky behaviours, protection reflects the use of protective devices, and response refers to reactions to warning signals indicating the presence of a hazard. The current study applies this

theoretical framework to analyse our data, although we did not conceive our experiment based on its principles. Since route choice preceded participants' encounter with warning cues, we regard it merely as contextual exposure information rather than behavioural consequences driven by warning signals.

The remainder of this paper begins with reviewing pertinent warning interface literature and stating the directional hypotheses that direct our analysis. Next, we explain our experimental design, including immersive VR scenario, warning interface manipulation, procedure, dependent variables, and statistical methods. Thereafter, we report the results on post-trigger speed regulation, repeated trials, route choice, elapsed time, completion status, and questionnaires. Finally, we discuss the implications of our findings in connection to existing studies, outline limitations to inferential validity, identify priorities for component-matched VR studies of imminent landslide warning designs, and conclude.

Background and exploratory expectations

Previous studies of dynamic visual warnings have shown a basis for expecting different behaviors due to the difference in the two warning formats. Sudden appearance of visual stimuli captures attention [16]; dynamic/animated warnings affect the detection of visual stimuli and judgment about time [10, 12]; studies regarding Active Rearview Mirror Head-Up Display (AR-HUD) interfaces reveal that augmented warnings impact take-over efficiency and gaze behavior in simulated driving scenarios [21]; finally, previous VR hazard communication studies reveal that depending upon the configuration of the warning messages, users' engagement in safety information varies under time constraints [9]. All these findings do not indicate that a single element within the current study's active beacon warning will cause an effect. However, they motivate us to test certain directional expectations regarding the behavioral effect of the overall warning treatment.

Since no preregistration took place prior to conducting analyses, all directional statements below represent exploratory expectations rather than confirmatory hypotheses. Firstly, exploratory expectation E1 posited that participants exposed to the active-cue warning treatment will display relatively lower mean speed from 0 to 10 seconds post combined cue-hazard triggers on Attempt 1 compared to those exposed to passive-cue treatment. Secondly, exploratory expectation E2 suggests that in case the effect lasted beyond the initial orienting phase, participants who received active cues would continue exhibiting relatively lower mean speed between 0 to 20 seconds and over two attempts.

In line with the above expectations, RQ1 investigates whether there exists a difference between participants' mean speed measured over the 0 to 10-second window following combined cue-hazard triggers in Attempt 1, depending upon their warning condition. On the other hand, RQ2 assesses whether this expected difference persists when measured over 20 seconds and two attempts. Finally, RQ3 examines the outcome of participants' decision-making choices regarding route selection, attempt success, timing measures, and their responses on questionnaires concerning task comprehension.

RQ3 intentionally remains descriptive in its design since route choice was decided before receiving cues and cannot be regarded either as evidence of warning ineffectiveness or effectiveness [17].

Methods

Design

In this exploratory experiment, the participants were exposed to either active or passive cues, based upon a completely randomized design. The participants were randomly assigned prior to the cue treatment using computer-generated randomization scheme. Specifically, for every participant, a random number r was generated in [0,1], and those whose number was 0.5 and greater belonged to group 1 while those who obtained less than 0.5 belong to group 2, representing either the active cue group or the passive cue group. The participants had no choice in choosing the group, nor did they know about the other groups' cues. Attempt (1 vs 2) was a within-participant variable. Route manipulation and balance was avoided since it was an observed exposure behavior.

The study protocol was reviewed and approved by the Institute Ethical Committee (Human), Indian Institute of Technology Mandi (approval reference IITM/IEC(H)/2025/VD/P2), which met on 24 March 2025. Approval was granted under the project "Immersive Virtual Reality Applications for Environmental Disaster Preparedness, Defense Training, and Indian Knowledge System Interventions" and is valid for the duration of that project; the work reported here falls within its scope. The study was conducted in accordance with the Declaration of Helsinki [18] and with institutional guidelines for research with human participants. Written informed consent was obtained from all participants prior to participation.

Variables and analytic roles

Random assignment to conditions was performed for the key independent variable of Cue Condition (active illuminated beacons vs. passive pictographic signs). Repeat Within Participants factor is attempt number (attempt 1 vs. attempt 2). Observation Exposure Variable is the route chosen by participants, while the key Dependent Variable measured is mean car speed (km/h) in 0–10 seconds following the trigger in attempt 1. Further, other secondary and exploratory dependent variables considered are mean speed in 0–20 seconds, speed at trigger, speed difference at trigger, completion status, trigger duration, route selection, and responses to questionnaires. Fixed Effects factors in sensitivity analysis include hazard encounter ordinal and routes, and the distinction between independent variables that were manipulated, repeat within-participants factor, dependent observation exposure variables, and outcome variables is crucial since cue condition alone was randomly assigned.

Participants

Thirty-five participants took part (17 active cue, 18 passive cue), each undertaking two driving attempts. Mean age was 20.8 years (*SD* = 1.0, range 19–24); 27 were men and eight were women. Seventeen reported less than one year of driving experience, seven reported one to three years, and 11 reported four to seven years. Few reported personal experience of landslides. The sample was a convenience

sample of undergraduate engineering students with normal or corrected-to-normal vision, recruited at the host institution, and should not be considered representative of licensed or professional mountain-road drivers. All participants gave written informed consent.

Virtual-reality environment

The simulation was built in Unity 6.4 using the Universal Render Pipeline [27] and depicted a rainy mountain environment. Two roads departed from a common intersection: a 3.0 km road containing three landslide zones and a 4.5 km road containing one. Rainfall, lighting, road structure, intersection signage, and the positions of hazard zones and cues were held constant across conditions. Hazard zones and cues occupied fixed, pre-authored positions in the build, and the environment contained no other vehicles or pedestrians; these are properties of the simulation build configuration rather than facts recoverable from the telemetry. Hazard events were triggered when the vehicle crossed a spatial activation boundary, so event timing depended on the participant's approach speed while event location did not. Participants completed a flat-terrain practice drive beforehand. Table 1 lists environment and instrumentation parameters. Figure 1 summarises the route architecture and event sequence; it is schematic and does not imply unreported spatial coordinates. As shown in Fig. 1, both routes originate at a common decision point at which no warning cue is yet visible, so route selection necessarily precedes cue exposure.

Table 1
Environment and instrumentation parameters. Route distances, zone counts, cue modes, sampling interval and frame rate were recovered from the log files; the remaining rows are properties of the simulation build configuration reported by the authors.

Parameter	Hazard-dense road	Longer road
Route distance (km)	3.0	4.5
Hazard zones	3	1
Warning cues (one per zone)	3	1
Median traversal time, completed runs (s)	344 (n = 2)	433 (n = 26)
Cue mode logged, active-cue condition	WarningPole	WarningPole
Cue mode logged, passive-cue condition	SignOnly	SignOnly
Other vehicles or pedestrians	None	None
Hazard and cue placement	Fixed, non-adaptive	Fixed, non-adaptive
Telemetry interval	Nominal 100 Hz; median observed interval 0.010 s (99.9% of positive intervals)	Same as hazard-dense road
Display frame rate	Per-file median 70.4–72.1 frames/s	Same as hazard-dense road
Headset	Meta Quest 3; 4K render resolution; 72 Hz refresh; wireless Air Link	Same as hazard-dense road
Steering wheel and pedals	Thrustmaster T300 GT wheel and pedal set	Same as hazard-dense road
Transmission	Automatic; no gearbox controller used	Same as hazard-dense road
Host workstation	AMD Threadripper PRO 7975WX CPU; NVIDIA GeForce RTX 4090 GPU; 128 GB RAM; 4 TB SSD	Same as hazard-dense road

Warning-cue manipulation

Active cues involved lighting up a warning pole during the occurrence of hazard events, whereas passive cues involved observing the presence of a static triangular symbol pictograph of landslides which were constantly displayed regardless of any hazard occurrence (Fig. 2). Both warning signals were deployed similarly within the same hazardous zone. Neither warning signal format shares identical characteristics of perception. Apart from difference in temporal dynamics, differences exist in low-level visual features,

such as shape, color, luminance level, semantics, and commonality. Hence, it is difficult to attribute the observed effects solely to temporal salience. Moreover, none of the two warning signal formats represent a standardized practice adopted by field practitioners. Practically speaking, the logged trigger value in active cues refers to the total duration encompassing the cue activation phase and hazard occurrence, since the telemetry record doesn't separate timestamps for cue-onset, hazard onset, and first hazard appearance. It follows that no warning-only latency period is determinable from current findings. As Fig. 2 illustrates, the two devices differ not only in whether they change state but also in shape, colour and luminance, which is why the comparison is treated as one between intact warning formats.

Procedure

Once obtained informed consent and demographics, participants undertook a practice drive followed by participation in the two-road environment simulation task. Upon reaching the intersection point in the scenario, the information signage showed two routes ahead, and the participants were allowed freedom of choice concerning route selection; neither any direction nor incentive was provided by the researchers about which route to select, and the participants were free to choose again during the second attempt if they wished to do so. Hazard zone warning cues were experienced only after route selection. During driving simulation, the participant was instructed to drive as normally as possible, maintaining speed control within limits. Moreover, participants did not receive any information about the purpose of warning cues presented throughout the scenario. Hardware utilized in the experiment included Meta Quest 3 headset running wireless Air Link connection at a render resolution of 4K and refresh setting of 72 Hz, coupled with Thrustmaster T300 GT steering wheel and pedals. The gearbox was kept automatic. Two attempts were required from each participant, and no route constraint was placed on them regarding the second attempt.

The simulation software automatically recorded each participant's route-choice event and subsequent occurrence of single-trigger hazard events at each respective hazard zone which corresponded to cue mode experience, landslide occurrence and classification. Unfortunately, due to limitation in data acquisition, separable timestamps regarding onset time of the beacon, beginning of the hazard condition, and time of its initial appearance cannot be extracted from the simulation log files. Post-trigger times therefore relate to the overall warning-hazard event experienced. The shorter hazard dense road (3.0 km) offered the potential contribution to three hazard encounters per participant, whereas the long hazardous road (4.5 km) offered one encounter at the later hazard zone if completed.

Measures

Telematics included data on position, rotation, velocity, engine revolutions per minute (RPM), transmission gear position, throttle position, braking pressure, steering position, handbraking, clutch position, driver assistance status, frame rate, and discrete triggers. Exposure was operationalized based on route taken, while behavioral responses were operationalized primarily based on vehicular speeds.

Post-warning-hazard speed was instantaneous speed at the telemetry sample closest to the warning-hazard trigger. For the purpose of data analyses, we defined primary end-point as mean speed during 0–10 seconds after the warning-hazard trigger within each attempt 1 trial and mean speed within 0–20 seconds as secondary temporal sensitivity end-point. In scenarios where multiple hazards occurred in an attempt, mean speed values for each encounter were averaged to derive one mean value for each individual for each attempt to avoid assuming independence in participant-level analysis. For completion, we recorded time elapsed as total recording time of the trial, while decision was operationally defined as a specific decision recorded in common intersections. Response latencies were calculated based on recordings of brakes and accelerators but not presented here due to instability resulting from telematic data structures (as will be discussed later in Results section). Post-trigger speed data represent the behavioral responses towards the combined warning-hazard condition instead of warning-only detection because the timestamps of warnings and first visibility of hazards are not available separately.

Statistical analysis

This paper does not report any preregistration of the study design. Thus, all analyses in this paper should be taken as exploratory (i.e., no hypothesis testing is claimed). The average participant level speed (km/h) in the 10-second period after the cue and hazard combination on Attempt 1 was considered the main endpoint for the hazard-reaction time comparison between conditions. The remaining endpoints are secondary or exploratory and are not corrected for multiple comparisons. Any alternative analyses targeting the same endpoint are discussed in this paper as robustness checks rather than additional independent findings. Two-tailed tests with $\alpha = .05$ significance levels were used throughout. Confidence intervals for mean difference effects were based on pooling within-group variances. Assumptions of homogeneity of variance were tested by Levene's test and assumption of within-group normality was tested using Shapiro-Wilk W-statistic tests. In the absence of clear indication of violations of those assumptions, the corresponding parametric test was treated as the primary result and supplemented by distribution-free robustness checks. While random assignment increases internal validity for comparisons of the two full warning format groups, effect sizes and confidence intervals become critically important since the main outcome variable was selected ex-post, and both sample size in the analysis and differences between the warnings on visual and semantic dimensions are quite substantial.

The distribution-free robustness checks included an exact Mann-Whitney U test, and permutation tests. For change-score contrasts, permutations were conducted across all possible partitions of the group membership variable. Otherwise, permutations involved Monte Carlo resampling of 20,000 replicates with the fixed random seed 42. For binomial completion, contributing two attempts each, logistic generalised estimating equations with exchangeable correlation were fit using participant clustering. For encounter level sensitivity analysis, linear mixed models accounting for all 54 encounters were fitted by restricted maximum likelihood estimation method using a participant specific random intercept and condition, attempt, ordinal encounter and route fixed effect predictors. Fitting required Powell solver due to convergence issues. Default quasi-Newton fitting failed to converge returning warnings of singular

covariance matrix and boundary problems. Power sensitivity analysis determined the smallest detectable effect size at 80% power for the observed number of observations per cell. Missing questionnaire responses were only missing two values for one item. Results without any imputations are presented here as primary results and median imputation was considered a sensitivity check procedure. Imputation of any telemetry-based outcome variables did not apply. Telemetry data were preprocessed using Python version 3.12.3 [28]. Aggregations and alignments to events were performed using pandas version 3.0.2 [22], and NumPy version 2.4.4 [23]. Hypothesis testing, estimation, confidence interval calculation and figure preparation was done using SciPy 1.17.1 [24], and statsmodels 0.14.6 [25], with figures prepared using Matplotlib 3.10.8 [26].

Results

The Results are organised around exploratory expectations E1 and E2 together with RQ1–RQ3. Because E1 and E2 were not preregistered, findings are described as consistent or inconsistent with these expectations rather than as confirmatory support or rejection. The participant-level 10 s comparison addresses E1 and RQ1; the 20 s and repeated-attempt analyses address E2 and RQ2; route choice, completion, elapsed time, and questionnaire measures address RQ3.

Participant flow and analytic denominators

Thirty-five participants (17 active cue, 18 passive cue) each undertook two driving attempts, giving 70 attempts and 70 telemetry files. A route-decision event was recorded on 65 attempts; five attempts ended without a recorded route-decision event. Paired route transitions were therefore available for 31 participants (16 active cue, 15 passive cue). Hazard-response analyses required a participant to reach and trigger at least one hazard zone: 23 did so on Attempt 1 (11 active cue, 12 passive cue) and 20 did so on both attempts (10 active cue, 10 passive cue). Across the dataset 54 hazard encounters were logged (28 active cue, 26 passive cue). Denominators differ across analyses because participants contributed only where the relevant event occurred, and this conditional structure substantially reduces effective sample size. Table 2 reports the denominator and unit of analysis for every result.

Table 2
Denominators, units of analysis, and tests for the principal telemetry outcomes. Questionnaire comparisons are reported in the text. E1 and E2 are exploratory expectations; no analysis plan was preregistered.

Outcome	Unit	n	Test	Result
Post-trigger speed, 10 s (primary)	Participant	11 vs 12	One-way ANOVA; Mann–Whitney; permutation	p = .041 / .046 / .039
Post-trigger speed, 20 s	Participant	11 vs 12	One-way ANOVA	p = .007
Post-trigger speed, all encounters	Encounter	54	Linear mixed model; clustered OLS	p = .026 / .028
Trigger speed, first encounter	Participant	11 vs 12	One-way ANOVA	p = .615
Change in trigger speed across attempts	Participant (paired)	10 vs 10	ANOVA on change scores; permutation	p = .023 / .019
Successful completion	Attempt (clustered)	34 vs 36	Logistic GEE, exchangeable	p = .086
Elapsed time, Attempt 2	Attempt	16 vs 16	One-way ANOVA	p = .009
Route choice	Participant	17 vs 16; 16 vs 16	Fisher exact	p = .656 / .394

Speed following the combined cue–hazard trigger (primary analysis)

Regarding Research Question 1 (RQ1), participants assigned to the active-cue condition drove relatively slow 10 s after the simultaneous occurrence of the cue and hazard stimulus, i.e., M = 21.6km/h, SD = 9.4, n = 11, compared to those who were assigned to the passive-cue condition, M = 32.6km/h, SD = 14.1, n = 12, $F(1, 21) = 4.74$, $p=.041$, $\eta^2=.184$. The mean difference (active minus passive) was − 11.0 km/h, with a 95% CI (pooled variance) of [− 21.5, − 0.5], and Cohen's d was equal to .91. Results obtained from Levene's test revealed no violations of the assumption of homogeneity of variances (p=.540); similarly, tests performed using the Shapiro-Wilk statistic revealed no violations of the assumptions of normality distribution (p=.750 and p=.545). Using an analysis of non-parametric data (i.e., distribution-free check), the Mann-Whitney U value was 33.0 (p=.046) while Monte Carlo permutations were 33 (p=.039). Both these statistical methods provide a measure of reliability for the analysis performed. Nonetheless, the estimation of the confidence interval barely exceeds zero. The random assignment process can be used only to make inferences about the differences observed among the two conditions (warning formats) but cannot draw conclusions on changes occurring in time. Finally, the direction of the results is supportive of the exploratory prediction E1. These participant-level means are plotted in Fig. 3, which shows the active-cue group sitting roughly 11 km/h lower than the passive-cue group, with error bars that overlap appreciably and so convey the imprecision of the estimate.

With respect to RQ2, results showed significantly reduced speeds in terms of the average speed during the 20s after trigger period in active-cue compared to the passive-cue condition (17.4Km/h vs 29.8km/h, F = 8.80, p=.007). As the windows 10s and 20s overlap each other, the latter represents the temporal sensitivity analysis and not the independent confirmation of the former. With respect to our exploratory analyses, we ran a Linear Mixed Effects Model (LMM) based on all 54 encounters and with random intercept of participant, along with fixed factors including conditions, attempts, encounter ordinal, and routes. The active-cue coefficient was − 7.73 km/h at 10 s (95% CI [− 14.54, − 0.91]), p=.026 and 9.71km/h at 20s, 95% CI [− 15.58, − 3.84], p=.001, relative to control condition. This analysis is sensitive as well since the encounter ordinal variable had extremely few cases (49 first encounters, three second, and two third); subsequent encounters occurred only in dense-hazard roads, and convergence needed Powell optimizers instead of default quasi-Newton method due to singular covariance warning. Using participant clustered Ordinary Least Square regression resulted in comparable estimates (− 7.13km/h, p=.028), although the statistical inference from 35 clusters must be interpreted with caution. Sustained effect over the 20s interval is consistently aligned with the exploratory hypothesis E2.

The average speeds did not show any significant differences in terms of combined cues and hazards triggers at the first encounter (49.1km/h vs 52.2km/h, mean difference 3.2km/h, 95% CI [− 9.9, 16.3], p=.615). However, the absence of any effects must be cautiously interpreted due to wide confidence intervals. Brake onset and pedal responses have been investigated. Although we do not report on these findings, the brake onset revealed conflicting ordering in relation to time censoring beyond 10s (out of total 11 and 12 participants under active and passive conditions, respectively, ten participants already pressed down the throttle pedal).

Change across attempts

In the current study, a group of 20 participants, out of the total 35, were those who triggered hazards twice in each condition (i.e., 10 in each condition). Their mean speed before triggering hazards decreased significantly from 50.80 to 48.57 km/h in the active-cue condition. Conversely, their mean speed before triggering hazards increased significantly from 51.73 to 59.23 km/h in the passive-cue condition (refer Fig. 4). An analysis of variance (ANOVA) based on the comparison between the two conditions showed that there exists a significant difference in mean changes between the conditions $F(1,18) = 6.21$; $p = .023$; the exact permutation $p = .019$; Exact Mann-Whitney $U = 22.0$; $p = .035$. The difference in mean changes is 9.72 km/h, the pooled variance 95% confidence interval (CI) is [1.52, 17.91]. This represents the minimum sample used in making inference regarding change of speed for hazard avoidance task. Figure 4 displays this crossover: the two conditions begin at almost the same speed on Attempt 1 and diverge on Attempt 2, the active-cue line falling while the passive-cue line rises.

Route choice, completion, and elapsed time

For RQ3, selection of a route precedes exposure to cues and thus, will be considered as elected exposure. Out of 31 participants with pairs of route choices, 25 chose longer roads twice, 5 chose the hazardous roads twice, and 1 changed his/her choice to hazardous roads from longer roads; no

participant changed his/her choice from hazardous roads to longer roads. There was no significant difference in choosing hazardous roads based on cues (Attempt 1: 2/17 vs. 3/16, Fisher exact p=.656; Attempt 2: 2/16 vs. 5/16, p=.394). As this study was not designed as an equivalence test, these non-significant findings cannot confirm the equivalence of route preferences between cue conditions. These findings simply describe the route choice context prior to the warning format comparison.

Out of 34 trials under active-cue, 18 cases were completed successfully and out of 36 trials under passive-cue, 10 cases were completed successfully. Since each participant provided two data points, the observations are not independent. By running a logistic GEE model with exchangeable working correlation matrix, we obtained odds ratio equal to 2.93 with 95% confidence interval [0.86, 9.96], p = .086; participant level comparison resulted in Mann-Whitney p = .071. The estimated odds ratio indicates better performance in active-cue, but since the confidence interval covers null value, success in completing the trial remains inconclusive among the subjects involved.

Mean time spent on second attempt was higher for participants under active cue (Mean = 397.6 sec, SD = 137.8) than those under passive-cue (Mean = 252.5 sec, SD = 153.9); F(1,30) = 7.90, p=.009. Time depends jointly on both route chosen and early termination. Since caution must be measured by modelling routes and early terminations together, time will be used here merely as task context information, but not as a safety outcome measure.

It was impossible to conduct a habituation test since two participants only encountered more than one hazard zone within attempts; contribution was 8 out of 54 total encounters, due to preference to single-hazard zone roads and preterm termination in hazardous zone choice scenarios.

Questionnaire measures

The pre-task variables were examined for any differences between conditions concerning mountain driving risk perception (M = 3.5 vs. 3.3; F(1,34) = 0.21; p=.44), landslide risk perception (M = 3.5 vs. 3.5; F(1,34) = 0.01; p=.93), pre-task anxiety (M = 2.2 vs. 2.8; F(1,34) = 0.69; p=.41), realisticness perception (M = 3.9 vs. 3.6; F(1,34) = 0.95; p=.34), comfortability rating (M = 3.8 vs. 3.3; F(1,34) = 1.07; p=.31), and driving experience (M = 4.0 vs. 3.6; F(1,34) = 0.98; p=.33). The confidence in navigation skills in the active-cue condition was numerically lower (M = 2.6 vs. 3.3), but the difference was not statistically significant (F(1,34) = 4.34; p=.053). These preliminary differences must be treated carefully since sample size was relatively small (n = 17 active-cue; n = 18 passive-cue). In addition, there were two missing values on one item – speed-preference. Using complete-case approach (n = 32), we found no differences (p=.75), however using median imputation method to deal with missingness provided similar results (p=.80).

Participants who took route with high density of hazards (n = 5) and took another route with fewer hazards (n = 27) were similar concerning seven out of eight risk-attitude measures (i.e., they had no statistically significant differences). The sole difference was found on arrival speed vs. safety question (M = 3.00 vs. 2.15; W = 79.00; p=.085) with only five participants opting for faster arrival; therefore such

difference is not conclusive due to low power. There might also be concerns about confounding with exposure level which seems likely.

Finally, post-task questions concerning recognition of cue and its impact on performance showed no evidence that active-cue participants recognized warning pole better than passive-cue participants (12/17 vs. 8/18, respectively). In fact, both cue clarity ratings (M = 3.71 vs. 3.33; F(1,34) = 1.75; p=.215) and cue impact on performance ratings (M = 3.65 vs. 3.44; F(1,34) = 0.53; p=.483) did not differ. It is not convincing enough whether this is good evidence for cue manipulation validity.

Discussion and conclusions

Our main finding was lower average speed after cue activation among participants randomly assigned to receive the active illuminated-beacon treatment compared with those randomly assigned to receive static pictographic signs. The results of the 20-second window sensitivity analysis corroborated the findings reported above. Furthermore, the results using the reduced sample size with repeated trials suggested even less pronounced changes in trigger speed for participants receiving the active cue. These findings are consistent with our exploratory predictions E1 and E2. On the other hand, the choice of route did not affect warning effectiveness because it occurred before cue presentation. Route completion, however, remains a problematic construct to evaluate because of repeated attempts, so our conclusions relate to differences in average speed only. Random allocation strengthens causal interpretation of differences between the assigned warning formats, although the exploratory analyses, reduced effective sample sizes and conditioning on hazard encounters limit the strength and generalisability of that inference. However, we would stress that we are only providing evidence for a difference between the two warning format levels and not temporal state change per se.

The conclusions made above are consistent with existing literature on human factors engineering. Studies on the impact of virtual reality-based hazard communication suggest that configuration matters for engaging safety messages [9]. Simulation studies on warning effects reveal varying improvements on detection, time perception, gaze, and takeover depending on context and desired outcome [10], [12], [21]. In fact, recent reviews highlight the promise but also limitations of immersive technologies in driver training and assistance [20]. In this study, the active warning format differs from the passive format not only temporally but also geometrically, chromatically, luminously, semantically, and contextually. Hence, low-level conspicuity, novelty, urgency, and temporal state change must compete as explanations. Our telemetry does not distinguish between beacon start, hazard initiation, and first visibility of hazards. The speed measurements therefore reflect the driver's reaction to the whole event consisting of both cues and hazards, not simply warning stimuli alone.

Some cautions are also required regarding inference in this study. Since the sensitivity analysis window overlaps with the initial window, only twenty participants participated in the repeated-attempts analysis, and route choices, completions, and speeds share participants within a common experimental paradigm, the apparent consistency of findings must be qualified. Multiple confidence intervals in this paper are

relatively large, and a sensitivity analysis suggests limited capacity for detecting moderate standardized effects. Therefore, null findings do not necessarily provide much information regarding small-sized effects. Finally, warning reliability was perfectly certain in this study compared with a real-world deployment, thereby reducing complexity. Our previous landslide hazard warning study demonstrates how reliability affects behavioural reactions [14]. False alarms reduce compliance in automation research [15]. Thus, it might be inappropriate to extrapolate findings into more realistic settings.

As discussed above, TriRB contributes to interpreting the results obtained here [17]. Participants decided to proceed regardless of cue exposure. Hence, route choice can indicate the willingness to accept risks but not warning effectiveness. Cue type is a feature of the interface under investigation, while speed after warning stimulus provides direct measurement of behaviour change induced by warnings. Such a distinction is necessary to prevent conflating pre-warning risk-taking with post-warning outcome. In this experiment, TriRB is valuable as an interpretive guide for determining which measures pertain to evaluating the behavioural interface.

Implications of our work suggest that it is necessary to assess roadside warnings' design based on driver interface characteristics rather than information displays solely. Telemetry will reveal driver reactions based on continuous measurement. Eye tracking and reaction times will identify when drivers perceive the warnings or hazards, so the combination of metrics improves reliability and validity [11]. Immersive research on road safety increasingly integrates both attention and behavioural measures to study interfaces [20], [21]. Specifically, research on landslide warning highlights that timing depends on observers' positions within scenes [7]. It is crucial to separate warning start times from hazard initiation and visibility.

The current limitations delineate specific areas requiring rigorous follow-up investigations. First, the sample size of the study was modest, homogenous, mostly comprising young engineering students with minimum driving experience. Hazard-response evaluations were conditioned upon successful encounter with hazards, thereby lowering effective samples for the main telemetry outcomes. Second, there was no preregistration of analysis plans; further, several endpoints were considered simultaneously without any multiplicity correction method; moreover, 10-second-endpoint became a primary outcome endpoint during post-hoc analyses. Other factors, such as lack of component matching among warning forms, reliability of warning, choice of self-selection of routes, missing timing points of warning initiation and visibility of the first hazard, absence of simulator-sickness measure, experience with gaming or simulators, and physiological measurements were also noted limitations. These limitations should not be overlooked but should be addressed as boundaries to infer mechanisms and generalization of claims to the real world. Accordingly, the observed between-format differences should be interpreted as exploratory and in light of these limitations.

It is recommended to begin with a component-matched comparison in terms of hazard warnings on mountain roads. Component matching entails comparing identical physical characteristics of warnings except for their transition behaviour, which needs to vary. Logging of various events is needed to

disentangle the effect of warning response from actual visibility of the hazard as well as to determine minimum speed. Specifically, the logged events will include: warning activation time, the time point when hazard becomes visible to a driver, time point of initiation of a hazard, throttle release, brake activation, and minimum speed attained before collision (i.e., hazard onset). Exposure to hazards must be standardized, so the amount of cues presented is independent from route selection. Warning reliability should be manipulated in realistic ranges to assess compliance, trust, and habituation due to repeated exposures [14], [15].

It is suggested to incorporate eye-tracking as an objective measure of warning cue detection and attention allocation in driving hazards. Eye-tracking is already proven in driving hazard perception research [11]. In addition to eye-tracking, measures should include assessment of simulator-sickness with validated tools [19] and history of simulation experience, both games-based and real-driving experience. Electrophysiological recording (EEG), as well as cardiovascular (heart-rate variability) measures might prove useful if cognitive workload or physiological stress are among the variables under examination in the experiment (it is required to specify those and their analysis plan ahead). Participants must be recruited not solely from college students' pool but also include experienced, licensed, and professional drivers familiar with mountain roads. Sample size estimation is to be done based on desired precision/power accounting for repeated-measures.

Translating the findings into practice requires verification of behavioural pattern across different conditions. Conditions include varying visibility levels, rainfall intensities, road geometry, and distances to warning activation points; additionally, experiments can compare findings with higher fidelity simulation and even field studies if feasible. Moreover, multimodal warning presentations might be examined since there is current trend towards immersive virtual reality studies into driving safety. Those multimodal warning systems include combinations of visual information presentation with audio cues and tactile sensation in order to assess driver performance in landslide-prone environments.

In an exploratory randomized experiment using virtual reality simulation of landslides, participants exhibited a significantly slower speed response in the case of an active warning format than the passive warning format while driving along two roads in a VR setting. While the increased post-trigger interval and repetition attempt showed consistency with the observed pattern, the choice route had no temporal relevance due to inability to capture warning efficiency. These exploratory results provide evidence that the two warning formats were associated with different speed-regulation responses in this virtual-driving environment. As warnings were non-component matched and there was no separate timestamp for warning onset from hazard occurrence, no conclusions about state change effect were made as the underlying cause of the difference between speeds. What this research contributes is a clear methodology of assessing interfaces in extended reality through randomization, telemetry measurement of driving actions, comparison of warning-format effects on drivers, and appropriate interpretation of results.

Declarations

Data availability

All necessary files for replication of the findings in this manuscript include the anonymized participant-level data analysis file, data dictionary, figure source data, and telemetry data without personal identifiers. These will be provided by the corresponding authors upon reasonable request. Participant identifiers, namely names, institutional roll number, email addresses, and the participant-identifier key, will not be disclosed due to concerns regarding participant confidentiality. All identifier data will remain confidentially stored within the institutionally controlled environment. Complete case analysis and median imputation analysis was conducted. None of the measures computed from the telemetry had missing values and did not require imputations. The authors will supply only necessary anonymized files required for editorial review.

Code availability

The analysis code used to produce all reported statistics and figures is available from the corresponding authors upon reasonable request. The archive comprises the analysis scripts and statistical software syntax, software and package versions, random seeds, exclusion and censoring rules, model specifications and diagnostics, and a table mapping each reported result to its source output.

Acknowledgements

This study was supported by the Tata Education Trust under the Tata SAFER Hills project (Grant ID 2025-73080), awarded to the Indian Institute of Technology Mandi, which provided the laboratory and computational resources used for this work. The authors thank all participants for their time and engagement.

Author Contributions

AM developed the virtual-reality simulation and experimental scenarios, conducted the experiment, performed the primary analyses and drafted the manuscript. KC assisted with data analysis. TG, JM, KVU and AKR provided guidance, contributed to interpretation of the findings and reviewed the manuscript. VD conceived and supervised the study and served as the principal guide. All authors reviewed and approved the final manuscript.

Additional Information

Competing interests

All authors declare no competing interests. Varun Dutt is a Guest Editor of the Scientific Reports Collection to which this manuscript is intended to be submitted and will take no part in the editorial handling or peer-review decision for this manuscript; editorial management should be assigned independently in accordance with journal policy.

Ethical approval

This study involved human participants and was conducted in accordance with the ethical standards and guidelines of the Indian Institute of Technology Mandi and with the Declaration of Helsinki [18]. The study protocol was reviewed and approved by the Institute Ethical Committee (Human), Indian Institute of Technology Mandi (approval reference IITM/IEC(H)/2025/VD/P2), which met on 24 March 2025. All participants provided informed consent prior to participation. No human tissue samples were involved.

Informed consent

Written informed consent was obtained from all subjects involved in the virtual-reality and questionnaire-based study.

Funding

This study was supported by the Tata Education Trust under the project “Tata Trust Sponsored Augmented Facilities for Empowering Resilience in Himalayan Hills (Tata SAFER Hills)”, Grant ID 2025-73080, awarded to the Indian Institute of Technology Mandi, which also provided the laboratory and computational resources used for this work.

References

1. Froude, M. J. & Petley, D. N. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* **18**, 2161–2181. https://doi.org/10.5194/nhess-18-2161-2018 (2018).
2. Intrieri, E., Gigli, G., Casagli, N. & Nadim, F. Landslide early warning system: toolbox and general concepts. *Nat. Hazards Earth Syst. Sci.* **13**, 85–90. https://doi.org/10.5194/nhess-13-85-2013 (2013).
3. Šakić Trogrlić, R. et al. Early warning systems and their role in disaster risk reduction. In Towards the ‘Perfect’ Weather Warning (ed. Golding, B.) 11–46 (Springer, 2022). https://doi.org/10.1007/978-3-030-98989-7_2
4. Endsley, M. R. & Jones, D. G. *Designing for Situation Awareness: An Approach to User-Centered Design* 3rd edn (CRC, 2025).
5. Zhu, Y. & Li, N. Virtual and augmented reality technologies for emergency management in the built environments: a state-of-the-art review. *J. Saf. Sci. Resil.* **2**, 1–10. https://doi.org/10.1016/j.jnlssr.2020.11.004 (2021).
6. Solinska-Nowak, A. et al. An overview of serious games for disaster risk management: prospects and limitations for informing actions to arrest increasing risk. *Int. J. Disaster Risk Reduct.* **31**, 1013–1029. https://doi.org/10.1016/j.ijdrr.2018.09.001 (2018).
7. Alene, G. H. et al. Realistic visualization of debris-flow type landslides through virtual reality. *Landslides* **20**, 13–23. https://doi.org/10.1007/s10346-022-01948-x (2023).
8. Buttussi, F. & Chittaro, L. Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE Trans. Vis. Comput. Graph.* **24**, 1063–1076.

https://doi.org/10.1109/TVCG.2017.2653117 (2018).

9. Cavalcanti, J., Valls, V., Contero, M. & Fonseca, D. Gamification and hazard communication in virtual reality: a qualitative study. *Sensors* **21**, 4663. https://doi.org/10.3390/s21144663 (2021).
10. Wang, Y. et al. Head-up display graphic warning to support collision avoidance: effect of graphic animation and border on driving behavior and eye movement pattern. *Transp. Res. Rec.* **2677**, 636–652. https://doi.org/10.1177/03611981221134628 (2023).
11. Cao, S. et al. Hazard perception in driving: a systematic literature review. *Transp. Res. Rec.* **2676**, 666–690. https://doi.org/10.1177/03611981221096666 (2022).
12. Zhai, W., Zhu, Q. & Hao, P. Dynamic visual warning cues and monochrome gradient types: key factors shaping AR-HUD warning interface design for drivers. *IEEE Access.* **13**, 195117–195131. https://doi.org/10.1109/ACCESS.2025.3632326 (2025).
13. Mehra, A. et al. Investigating behavioral responses across landslide scenarios in virtual reality. In Proceedings of the 10th International Congress on Information and Communication Technology 103–114 (Springer, 2025). https://doi.org/10.1007/978-981-96-6429-0_9
14. Mehra, A. et al. Virtual reality in crisis: how perceived risk and warning accuracy shape behavior in landslide scenarios. In 2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops 687–694 (IEEE, 2025). https://doi.org/10.1109/VRW66409.2025.00140
15. Wickens, C. D. & Dixon, S. R. The benefits of imperfect diagnostic automation: a synthesis of the literature. *Theoretical Issues Ergon. Sci.* **8**, 201–212. https://doi.org/10.1080/14639220500370105 (2007).
16. Yantis, S. & Jonides, J. Abrupt visual onsets and selective attention: evidence from visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 601–621. https://doi.org/10.1037/0096-1523.10.5.601 (1984).
17. Ben-Asher, N. & Meyer, J. The Triad of Risk-Related Behaviors (TriRB): a three-dimensional model of cyber risk taking. *Hum. Factors*. **60**, 1163–1178. https://doi.org/10.1177/0018720818783953 (2018).
18. World Medical Association. WMA Declaration of Helsinki: ethical principles for medical research involving human subjects. (2024). https://www.wma.net/policies-post/wma-declaration-of-helsinki/
19. Kennedy, R. S., Lane, N. E., Berbaum, K. S. & Lilienthal, M. G. Simulator Sickness Questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* **3**, 203–220. https://doi.org/10.1207/s15327108ijap0303_3 (1993).
20. He, P., Bremner, P. & Booth, C. A. Immersive technology for road safety: a systematic review of empirical studies. *Virtual Real.* **30**, 41. https://doi.org/10.1007/s10055-025-01295-x (2026).
21. Jing, C., Shang, C., Yu, D., Chen, Y. & Zhi, J. The impact of different AR-HUD virtual warning interfaces on the takeover performance and visual characteristics of autonomous vehicles. *Traffic Inj. Prev.* **23**, 277–282. https://doi.org/10.1080/15389588.2022.2055752 (2022).
22. McKinney, W. Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference 56–61 (2010). https://doi.org/10.25080/Majora-92bf1922-00a

23. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362. https://doi.org/10.1038/s41586-020-2649-2 (2020).
24. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*. **17**, 261–272. https://doi.org/10.1038/s41592-019-0686-2 (2020).
25. Seabold, S. & Perktold, J. Statsmodels: econometric and statistical modeling with Python. In Proceedings of the 9th Python in Science Conference (2010).
26. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95. https://doi.org/10.1109/MCSE.2007.55 (2007).
27. Unity Technologies. Unity 6 User Manual. (2026). https://docs.unity3d.com/
28. Python Software Foundation. Python 3.12 documentation. (2024). https://docs.python.org/3.12/

Figures

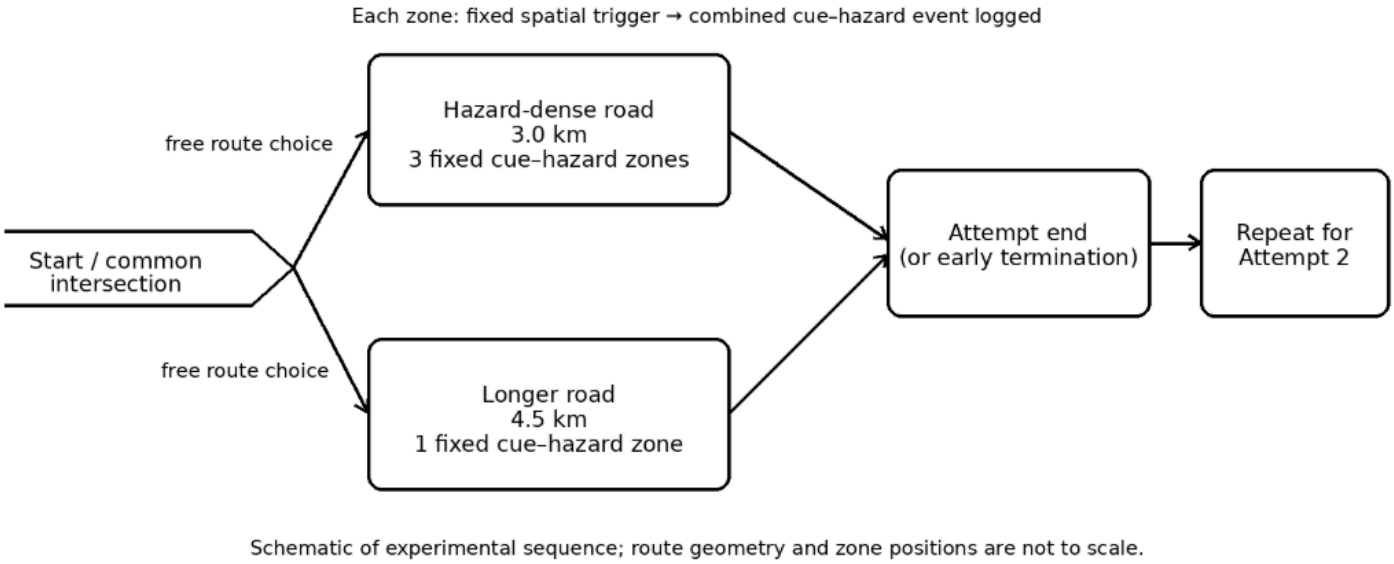


Figure 1

Schematic of the experimental sequence and route architecture. Participants chose freely between a 3.0 km hazard-dense road containing three fixed cue–hazard zones and a 4.5 km longer road containing one fixed cue–hazard zone. Each participant completed two attempts and could repeat or change route. Hazard events were triggered at fixed spatial boundaries; the schematic is not to scale.

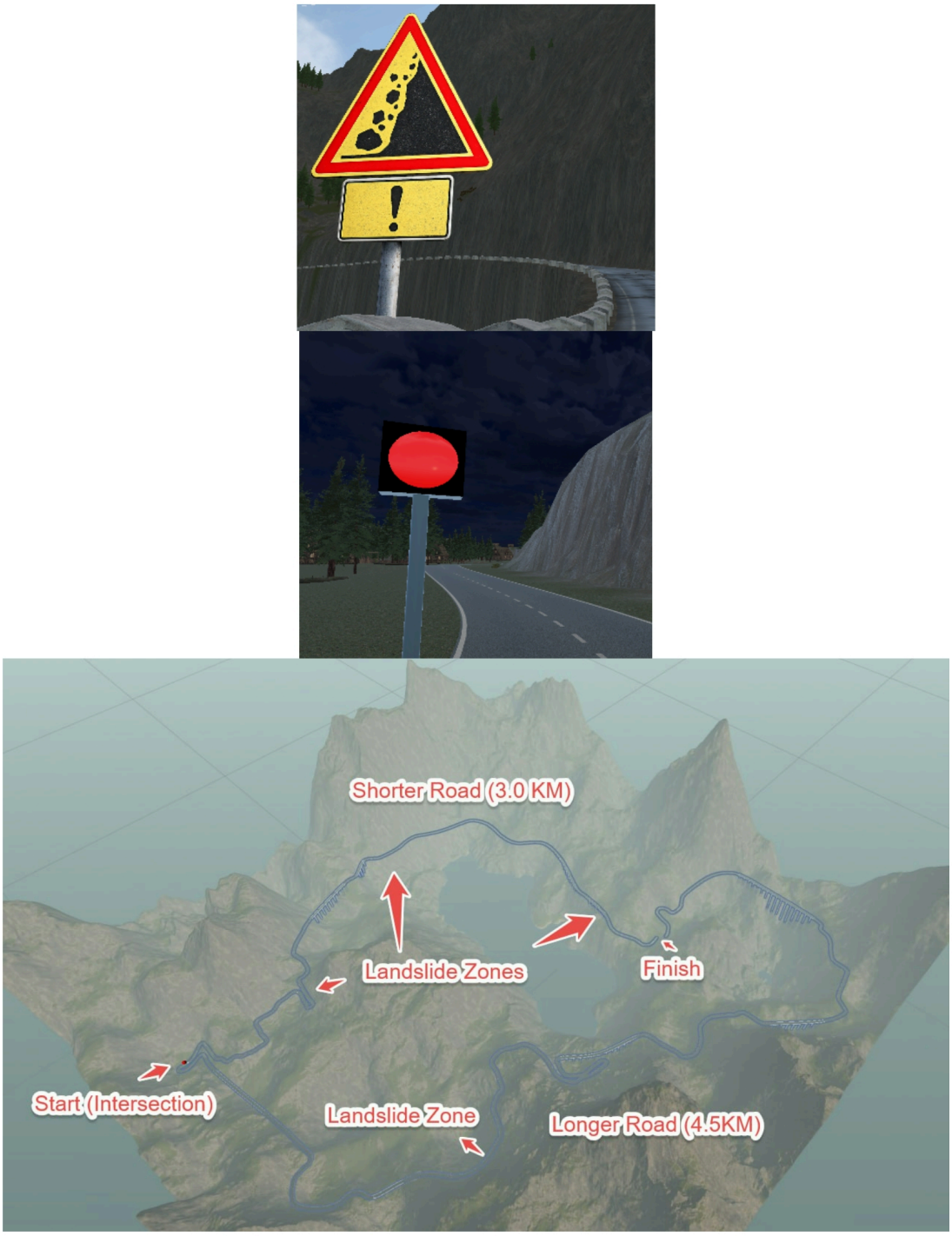


Figure 2

The two warning formats. Top: static pictographic landslide sign. Bottom: illuminated warning pole in the activated state. The devices differ in form, colour and luminance as well as in temporal behaviour. Lastly, we have a bird's eye view of the simulation.

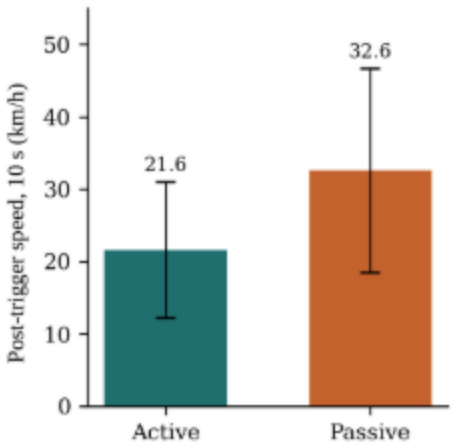


Figure 3

Mean speed in the 10 s following the combined cue–hazard trigger on Attempt 1, aggregated to one value per participant. Error bars denote one standard deviation. Active cue n = 11, passive cue n = 12.

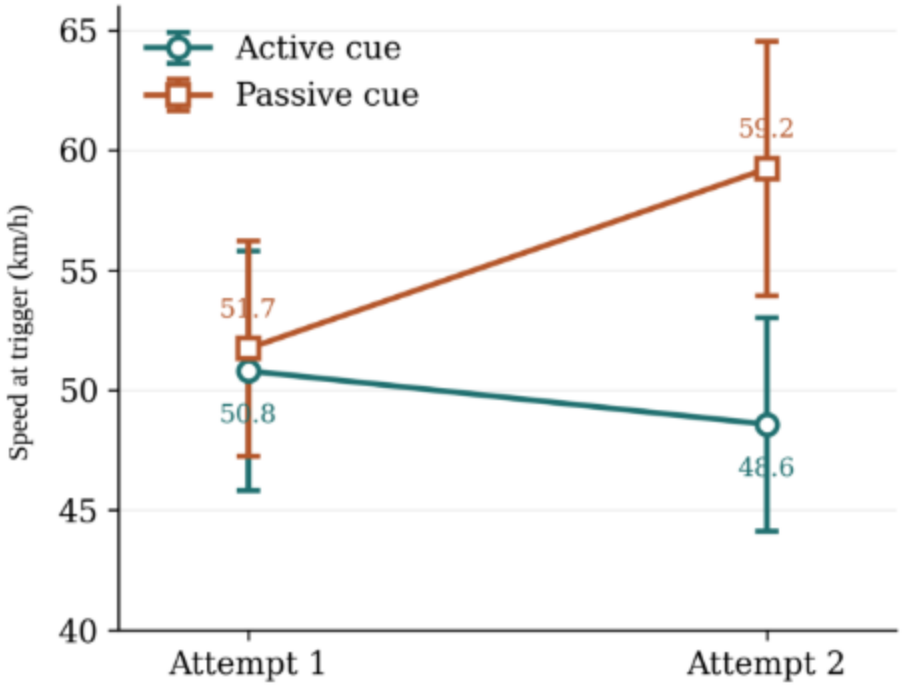


Figure 4

Mean speed at the combined cue–hazard trigger across attempts, for the 20 participants who triggered a hazard on both attempts (10 per condition). Points are group means of participant-level values averaged across all encounters within an attempt; error bars denote one standard error. Source data are available from the corresponding authors on reasonable request.