

Crosswalk Task: A naturalistic reinforcement learning paradigm for characterizing real-world risky behavior

Jinwoo Jeong

Seoul National University <https://orcid.org/0009-0001-1407-0179>

Woo-Young Ahn

wahn55@snu.ac.kr

Seoul National University <https://orcid.org/0000-0002-5900-8432>

Research Article

Keywords: risk taking, naturalistic reinforcement learning, inverse reinforcement learning, deep learning, pedestrian behavior

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11343272/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Crosswalk Task: A naturalistic reinforcement learning paradigm for characterizing real-world risky behavior

Jinwoo Jeong^{1,2} & Woo-Young Ahn^{1,3,4}

¹Department of Psychology, Seoul National University, Seoul, Republic of Korea

²School of Computer Science, Georgia Institute of Technology, Atlanta, Georgia, United States

³Department of Brain and Cognitive Sciences, Seoul National University, Seoul, Republic of Korea

⁴AI Institute, Seoul National University, Seoul, Republic of Korea

Corresponding author

Woo-Young Ahn

Department of Psychology, Seoul National University

Seoul, Republic of Korea

Telephone: +82-2-880-2538

Email: wahn55@snu.ac.kr

Abstract

Conventional laboratory tasks often fail to capture the dynamic, context-dependent nature of real-world risky behavior. To address this challenge, we developed the Crosswalk Task, a naturalistic reinforcement learning paradigm for measuring risk taking based on real-time behavior. Behavioral data collected from 309 adults were used to identify multiple task measures associated with trait-level risk preference, including aggregate behavioral measures and those derived using inverse reinforcement learning (IRL). Among these, IRL-derived measures captured context-specific subjective rewards inferred from state-action trajectories and provided more information about risk preference than aggregate behavioral measures. Subsequent regression and clustering analyses further demonstrated that jointly considering multiple Crosswalk Task measures can reveal information about individual differences in risk preference beyond that captured by individual measures. The Crosswalk Task thus provides a foundation for developing flexible, naturalistic reinforcement learning paradigms for studying decision making. *CrosswalkEnv* and *CrosswalkWebEnv*, the open-source software for task implementation and data collection, are publicly available on GitHub.

Keywords: risk taking, naturalistic reinforcement learning, inverse reinforcement learning, deep learning, pedestrian behavior

Introduction

Reinforcement learning (RL) has become a prominent framework for understanding human decision making by characterizing how individuals learn from experience and select actions to maximize long-term reward (Niv, 2009; Sutton and Barto, 1998). However, experimental research on human decision making typically relies on highly controlled and simplified tasks, raising questions about whether findings obtained in the laboratory generalize to behavior in more complex real-world settings (Cross et al., 2021; Nastase et al., 2020; Wise et al.,

2024).

Naturalistic RL paradigms seek to bridge this gap by incorporating elements of real-world complexity into controlled decision-making environments (Allen et al., 2024; Cross et al., 2021; Wise et al., 2024). This added complexity may enhance ecological realism, allowing trait-level psychological tendencies that shape real-world behavior to be more clearly reflected in decision-making processes within such tasks. For example, building on evidence linking impulsivity to reckless driving, a recent study found that measures derived from a real-time driving task were more strongly associated with self-reported impulsivity than those derived from conventional behavioral tasks (Hatfield et al., 2017; Lee et al., 2024). Naturalistic paradigms may therefore complement conventional behavioral tasks, whose measures often correlate weakly with self-report measures intended to assess the same construct (Dang et al., 2020).

Risk taking, in particular, may benefit from such an approach. Since the early days of risk-taking research, behavioral measures have been criticized for showing limited convergence across tasks (Slovic, 1962, 1964). Despite substantial methodological advances, behavioral measures of risk taking continue to show limited convergence with one another and with self-report measures, raising questions about whether these measures capture the same underlying construct (Fox and Tannenbaum, 2011; Frey et al., 2017). Recent meta-analytic evidence further suggests that the reliability of risk preference measures varies widely, with behavioral measures being particularly unreliable over time and across tasks (Bagaiini et al., 2025). These inconsistencies may reflect the context-dependent nature of risk taking. Accordingly, reviews have called for more realistic, externally valid tasks that better capture the dynamic expression of risk taking (Bran and Vaidis, 2020; Schonberg et al., 2011).

Pedestrian road crossing may provide a particularly suitable naturalistic paradigm for studying risk taking because it requires individuals to make decisions under uncertainty while facing the possibility of collision, injury, or death (Piyalungka et al., 2025; Theofilatos et al., 2021). When crossing roads, pedestrians must assess approaching traffic, evaluate

gaps between vehicles, and adapt their behavior based on changing conditions and prior experience (Cœugnet et al., 2019; Tian et al., 2023; Yannis et al., 2013). Motivated largely by efforts to reduce pedestrian collisions and fatalities, prior research has examined road-crossing behavior and its psychological correlates through observational, self-report, and experimental approaches (Aghabayk et al., 2021; Cœugnet et al., 2019; Hashemiparast et al., 2020; Liu and Tung, 2014). Recent studies have also shown that pedestrian road-crossing behavior can be modeled using RL (Wang, Srinivasan, et al., 2023; Wang et al., 2025), suggesting that road crossing may serve as a naturalistic reinforcement learning paradigm.

An important methodological consideration for naturalistic reinforcement learning paradigms is how to analyze the dense state-action trajectories generated in complex and changing environments. Unlike conventional reinforcement learning, which learns a behavioral policy based on a given reward function, inverse reinforcement learning (IRL) infers a latent reward function from observed behavior without requiring directly observed rewards (Arora and Doshi, 2021; Ng, Russell, et al., 2000; Ziebart et al., 2008). This approach allows the subjective values associated with different actions to be examined across environmental conditions. Maximum entropy deep IRL integrates deep neural networks (DNNs) into the maximum entropy framework to approximate reward functions, enabling complex, nonlinear relationships and interactions among state features to be captured without imposing a restrictive predefined functional form (Wulfmeier et al., 2015). Once learned, the inferred reward function can be used to estimate rewards for specific states or state-action pairs, providing a model-based approximation of individuals’ internal valuations or preferences. Previous studies have applied deep IRL to characterize individual differences in psychological traits from real-time behavior (Lee et al., 2024, 2026) and to model pedestrian crossing decisions using naturalistic traffic trajectories (Wang, Niu, et al., 2023). These applications illustrate the potential of IRL for analyzing complex behavioral data collected through naturalistic RL paradigms.

Taken together, these considerations motivated the development of the Crosswalk

Task, a naturalistic reinforcement learning paradigm for characterizing individual differences in risk taking. Across a range of simulated traffic conditions, the task requires participants to repeatedly choose between risk-taking and risk-averse strategies, balancing rapid progress against collision risk. To enable controlled laboratory experiments as well as scalable online data collection, the task is implemented in both desktop and web-based versions. Using behavioral data collected from more than 300 participants, we first defined and evaluated multiple task-derived measures as indicators of risk preference. We then applied IRL to the recorded state-action trajectories to derive participant-specific measures of subjective reward under distinct conditions. Finally, we used machine learning analyses to examine whether integrating multiple Crosswalk Task measures could provide a more informative characterization of individual differences in risk taking than any single measure.

To support the use and extension of the Crosswalk Task in behavioral and neuroscientific research, we provide open-source software implemented in Python and JavaScript. Researchers can use the software to conduct behavioral experiments, modify task configurations, and train or evaluate RL agents. By making the implementations publicly available, we aim to provide researchers with practical and extensible tools for using the Crosswalk Task and developing new studies of decision making in a dynamic, naturalistic environment. Installation instructions, usage examples, API documentation, and links to the open-source GitHub repositories are available in the official documentation (see Software below).

Crosswalk Task

Task Structure

The Crosswalk Task is designed to elicit risk-taking behavior that may reflect participants’ trait-level risk-taking propensity. In each episode, participants guide the player through a series of roads with the goal of reaching the top of the map as quickly as possible while avoiding collisions with cars (see Fig. 1). The player begins at the bottom of the map and



Figure 1. Screenshot of the Crosswalk Task. Participants guide the player across a series of roads as quickly as possible while avoiding collisions with cars. The current score increases after each successful road crossing. At the end of each episode, the best score (highest episode score) and total score (cumulative session score) are updated.

can move up, down, left, or right using the arrow keys or WASD keys on the keyboard, or remain stationary and wait. Participants must therefore continuously decide whether, when, and where to move while progressing toward the destination.

The task rewards both forward progress and rapid completion. Each time the player successfully crosses a road, the basic score increases in proportion to the vertical distance advanced. If the player successfully reaches the top of the map within the allotted time (e.g., 30 seconds), an additional bonus is awarded in proportion to the time remaining (see Fig. 2a). Thus, progressing farther through successful road crossings increases the basic score, whereas reaching the destination more quickly yields a larger bonus, creating an incentive to balance rapid advancement against the risk of collision.

Each episode can terminate in one of three ways. First, the episode ends successfully when the player reaches the top of the map, and a bonus score screen is displayed (see Fig. 2a). Second, a *Time Over* occurs if the player fails to reach the destination within the time limit. The remaining time is displayed continuously on the screen, with an optional color change to red when it falls below a threshold (see Fig. 2b). A *Time Over* ends the episode without any additional penalty or bonus; participants retain the basic score accumulated up

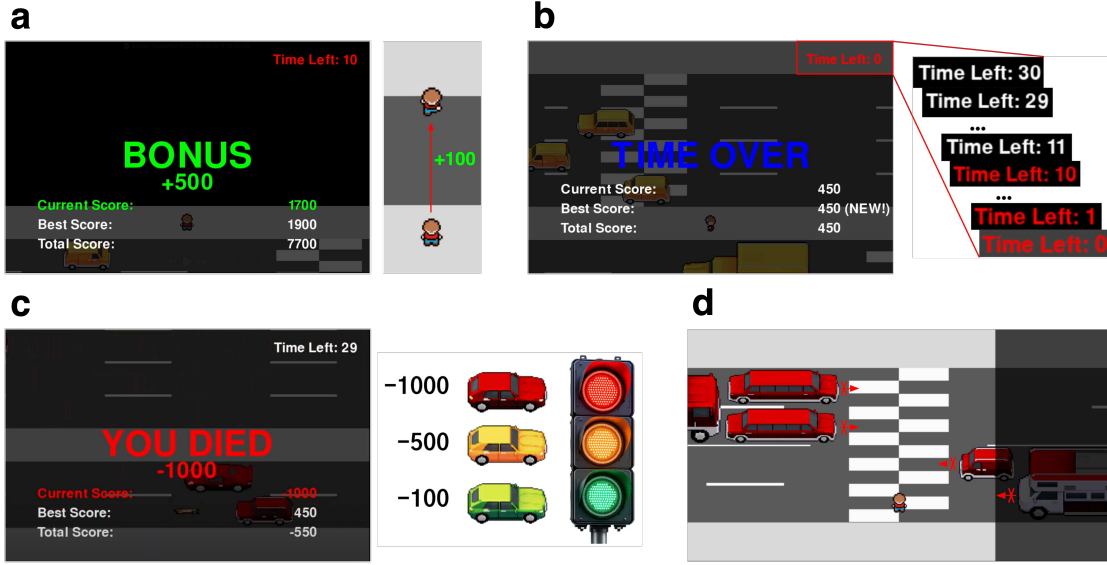


Figure 2. Key components and episode outcomes of the Crosswalk Task, illustrated using default values for configurable task parameters. (a) Bonus score screen and basic score increase following a successful road crossing, with default reward rates of 50 points per grid unit for the basic score and 50 points per second for the bonus score. (b) *Time Over* screen and changes in the remaining-time display, with a default time limit of 30 seconds and a red warning when fewer than 10 seconds remain. (c) *Run Over* screen and default collision penalties corresponding to traffic-light colors. (d) Crosswalk activation screen, in which approaching vehicles stop and wait for the player to cross safely.

to that point. Third, a *Run Over* occurs when the player’s hitbox makes contact with that of a vehicle, immediately terminating the episode and imposing a collision penalty.

All vehicles on a given road share one of three traffic-light colors corresponding to different penalty magnitudes: green, yellow, and red, representing low, medium, and high collision penalties, respectively (see Fig. 2c). Consequently, roads differ in the potential cost of a collision. Multiple penalty levels were intended to assess participants’ sensitivity to potential loss magnitude (Figner and Weber, 2011; Figner et al., 2009).

The key element of the Crosswalk Task is the crosswalk. Each road can contain at most one crosswalk, and some roads contain no crosswalk. A crosswalk is activated when the player stands immediately in front of the crosswalk or within it. Once the crosswalk is activated, approaching vehicles stop before entering it and wait for the player to cross (see Fig. 2d). Crosswalks therefore provide a substantially safer means of crossing a road,

but locating and activating them can require additional horizontal movement and time, thereby reducing the risk of a *Run Over* while increasing the risk of a *Time Over*. On each road, participants must choose between relatively risk-averse strategies, such as spending time searching for a crosswalk when none is visible or approaching a visible crosswalk, and relatively risk-taking strategies, such as jaywalking directly across the road even when a crosswalk is available. This recurring choice creates a safety–speed tradeoff: risk-averse strategies can eliminate collision risk through crosswalk use, whereas risk-taking strategies provide greater opportunities to complete the task faster and earn a larger bonus.

Participants repeatedly complete episodes during each experimental session, which may be defined by a fixed duration (e.g., 20 minutes) or a fixed number of episodes (e.g., 50). The traffic environment is procedurally varied both within and across episodes. Individual roads can differ in the number of lanes, the presence and location of crosswalks, and the size, width, and speed of cars, creating road-specific differences in difficulty and risk. The combination of these road conditions also varies across episodes, requiring participants to adapt their crossing strategies to the current conditions rather than rely on previously encountered patterns. Despite this procedural variation, the overall goal, control scheme, scoring system, and underlying risk–reward structure remain constant across episodes within a given experimental session. The variation in traffic conditions allows participants’ strategy choices to be observed across a range of situations, thereby providing a more comprehensive assessment of individual differences in risk-taking propensity.

Software

We provide two software implementations of the Crosswalk Task: *CrosswalkEnv* and *CrosswalkWebEnv*. Both implementations follow the same underlying task structure and are designed to produce compatible state–action trajectories while serving complementary purposes. *CrosswalkEnv* is a Python implementation built on the Gymnasium framework for running behavioral experiments and training or evaluating RL agents, whereas *Crosswalk-*

WebEnv is a web-based implementation for deploying the task in web browsers and collecting behavioral data at scale.

The two implementations support a common workflow spanning behavioral data collection and learning-based analysis. Behavioral trajectories collected with either implementation are represented in a compatible format, allowing data from *CrosswalkWebEnv* to be used with *CrosswalkEnv* for downstream analyses, including imitation learning and IRL. Further details on both packages, including installation instructions, usage examples, API documentation, and links to the GitHub repositories, are available in the official documentation (<https://crosswalk-task.github.io>).

CrosswalkEnv

CrosswalkEnv is distributed through the Python Package Index (PyPI) as `crosswalk-env`, and its latest stable version can be installed using `pip`:

```
pip install crosswalk-env
```

The package provides a high-level API for running the Crosswalk Task with a single function call. Experiments are configured and launched at the session level, with each session consisting of multiple consecutive episodes played without a break. For example, a 20-minute experimental session can be launched using the Python API as follows:

```
from crosswalk_env import task
task.run_session(subj_id=1, session_id=1, max_seconds=1200)
```

The same session can also be launched using the command-line interface (CLI):

```
python -m crosswalk_env.task --subj-id=1 --session-id=1 --max-seconds=1200
```

Both the Python API and CLI support additional task parameters for customizing the experimental settings. For example, the CLI command above is equivalent to the following command, which explicitly specifies task parameters that would otherwise use their default values:

```
python -m crosswalk_env.task \
  --subj-id=1 \
  --session-id=1 \
  --max-seconds=1200 \
  --episode-duration-sec=30 \
  --time-left-warning-threshold-sec=10 \
  --basic-score-per-grid=50 \
  --bonus-score-per-sec=50 \
  --red-car-penalty=1000 \
  --yellow-car-penalty=500 \
  --green-car-penalty=100
```

By default, each episode has a time limit of 30 seconds, with the remaining time highlighted in red during the final 10 seconds. These settings can be adjusted using `--episode-duration-sec` and `--time-left-warning-threshold-sec`, respectively. The warning can also be disabled by setting the latter to 0.

Scoring rules can be customized using `--basic-score-per-grid` and `--bonus-score-per-sec`, which determine the points awarded for each unit of upward movement and each second of remaining time, respectively. Similarly, collision penalties for red, yellow, and green vehicles can be configured independently using `--red-car-penalty`, `--yellow-car-penalty`, and `--green-car-penalty`. Setting any of these penalties to 0 disables the corresponding vehicle type from appearing.

The package also provides the `CrosswalkEnv` class, which implements the task as a Gymnasium environment. The class can be imported directly to run the task using modified environment configurations or to train and evaluate agents using standard RL algorithms.

CrosswalkWebEnv

CrosswalkWebEnv is distributed as the npm package `crosswalk-web-env` and provides the `ExperimentPage` React component for configuring and running Crosswalk Task sessions in a web browser. A single deployed website can support multiple participants and experimental sessions through participant- and session-specific configuration values passed to the component. A minimal example of a React component that runs a single 20-minute session is

shown below:

```
import ExperimentPage from "crosswalk-web-env";

export default function MainSession1({ subjId }) {
  return (
    <ExperimentPage subjectId={subjId} sessionId={1} maxSeconds={1200} />
  );
}
```

In a complete web-based experiment, participant identifiers and session configurations can be provided dynamically for different participants and sessions. For example, a participant identifier may be obtained from the URL, while different task configurations may be assigned to different URL paths. Detailed examples of how to implement such a website for conducting experiments are available in the official documentation (https://crosswalk-task.github.io/crosswalkwebenv/experiment_setup.html).

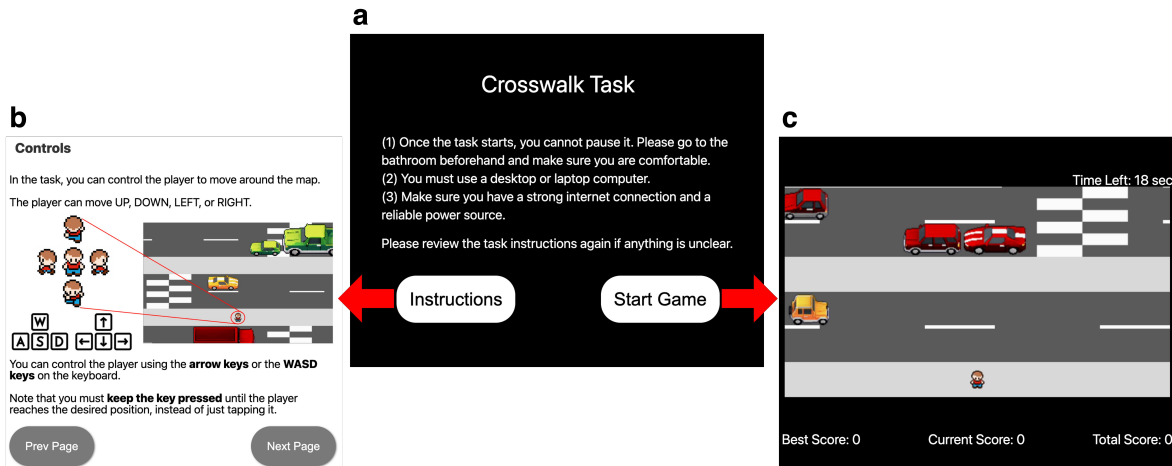


Figure 3. Example web pages rendered by the `ExperimentPage` component from the `crosswalk-web-env` package. (a) Main page presented before the session, with buttons for reviewing the instructions or starting the task. (b) Instructions page displayed after selecting the *Instructions* button. (c) Crosswalk Task screen displayed after selecting *Start Game*.

Figure 3 illustrates the interface rendered by the `ExperimentPage` component. Before starting a session, participants are presented with a main page from which they can review the task instructions or begin the task. Selecting the *Instructions* button opens the instructions page, whereas selecting *Start Game* starts the configured Crosswalk Task session.

Task Measures

A wide range of task measures can be derived from the Crosswalk Task, allowing researchers to select the measures most appropriate for their specific research questions. In this study, we categorize these measures into two broad classes: behavioral measures and IRL reward measures.

Behavioral Measures

Participants’ performance and decision making during the Crosswalk Task can be summarized using a variety of behavioral measures. These measures can be defined either as cumulative measures (e.g., *Total Score*) or as normalized measures based on averages or proportions (e.g., *Average Score*, *Episode Clear Ratio*, and *Run Over Ratio*). In this study, we focus on normalized behavioral measures because they are less affected by differences in task duration and therefore enable more meaningful comparisons across participants, experimental sessions, and studies.

Although these behavioral measures provide concise summaries of observed behavior, they do not explicitly characterize the latent decision processes underlying that behavior. Capturing such processes requires representations that preserve the sequential and context-dependent nature of decision making throughout the task.

IRL Reward Measures

Participant-specific reward functions were inferred using adversarial inverse reinforcement learning (AIRL; Fu et al., 2018; see Supplementary Information for details). From these functions, we derived reward measures for specific state–action conditions relevant to risk-taking decisions. Specifically, the inferred reward for moving up to enter a road without activating a visible crosswalk represents the modeled valuation of jaywalking, whereas the reward for moving sideways to activate a nearby crosswalk before crossing represents the

modeled valuation of crosswalk activation. These measures capture context-dependent valuations of specific decisions, complementing aggregate behavioral measures of overall task performance.

Data Collection

Participants

To evaluate the proposed Crosswalk Task, behavioral data were collected from an initial sample of 309 adults. The study comprised a desktop sample assessed in a laboratory setting and a web sample assessed online. Following data screening, a total of 282 participants were retained for analysis. Details of the data screening procedure are provided in the Supplementary Information.

Desktop sample Fifty-five Korean adults were recruited at Seoul National University. The desktop-based Crosswalk Task was administered in a laboratory setting following approval by the Institutional Review Board (IRB No. 2509/002-005). Following data screening, the final sample consisted of 53 participants (30 females, 23 males), with a mean age of 24.8 years ($SD = 4.9$).

Web sample We recruited 254 U.S. adults through Prolific. The web-based Crosswalk Task was administered online following approval by the Institutional Review Board (IRB No. 2603/003-037). Following data screening, the final sample consisted of 229 participants (121 females, 101 males, and 7 participants who preferred not to disclose their gender), with a mean age of 36.9 years ($SD = 10.7$).

Procedures

Questionnaires Participants completed the 30-item Domain-Specific Risk-Taking Scale (DOSPERT; Blais and Weber, 2006), which measures self-reported willingness to engage in risky behaviors across five domains: ethical, financial, health/safety, recreational, and social. For the web sample, responses were recorded on a seven-point scale ranging from 1 (*Extremely unlikely*) to 7 (*Extremely likely*). Korean participants completed the Korean version of the DOSPERT (Hong et al., 2010), which employs a five-point response scale. To account for the difference in response scales, total DOSPERT scores were normalized using the Percent of Maximum Possible (POMP; Cohen et al., 1999) method.

Participants also completed the Barratt Impulsiveness Scale (BIS-11; Patton et al., 1995) to assess trait impulsivity and the State–Trait Anxiety Inventory (STAI; Spielberger et al., 1983) to assess anxiety. Korean participants completed validated Korean versions of the BIS-11 (Heo et al., 2012) and STAI (Hahn et al., 1996).

Crosswalk Task Before starting the experiment, participants received an explanation of the Crosswalk Task and completed a 10-question comprehension quiz to ensure they understood the task. Participants who answered any questions incorrectly were required to answer those questions again until all answers were correct.

The desktop sample completed a desktop-based Crosswalk Task, whereas the web sample completed a web-based Crosswalk Task. The two versions differed only in their implementation (i.e., the programming language and software libraries used), whereas the task design, procedure, control scheme, and experimental conditions were identical.

Participants first completed a 5-minute practice session to familiarize themselves with the task, followed by two 20-minute experimental sessions separated by a 5-minute break. Except for session duration, all configurable task parameters were maintained at the same default values across both samples and throughout all sessions.

Behavioral Results

Before analyzing participants’ state–action trajectories in detail, we examined six behavioral measures to provide an overview of their task performance and behavioral tendencies (Fig. 4). These measures cover three aspects of task behavior: overall task performance (*Total Score* and *Average Score*), road-crossing performance (*Road Crossing Success Ratio* and *Episode Clear Ratio*), and risk-related behavior (*Jaywalking Ratio* and *Run Over Ratio*).

The *Total Score* represents the cumulative score earned across all episodes ($M = 98,171.809$, $SD = 22,698.480$), whereas the *Average Score* normalizes this value by the number of episodes completed by each participant ($M = 877.360$, $SD = 187.374$). Both measures exhibited approximately bell-shaped distributions, despite differences in the number of episodes completed across participants.

Regarding road-crossing performance, participants generally achieved high success rates at individual road crossings (*Road Crossing Success Ratio*: $M = 0.929$, $SD = 0.022$), whereas their success rates in completing entire episodes within the time limit were lower and more variable (*Episode Clear Ratio*: $M = 0.535$, $SD = 0.121$). The former exhibited a relatively symmetric distribution, while the latter showed greater variability, suggesting that both measures may be useful for examining individual differences in road-crossing performance.

Regarding risk-related behavior, approximately half of the roads with a visible crosswalk were crossed without using it (*Jaywalking Ratio*: $M = 0.508$, $SD = 0.161$). Its right-skewed distribution indicates that most participants exhibited moderate jaywalking ratios, with progressively fewer participants exhibiting higher ratios. Meanwhile, approximately 42% of episodes ended in a vehicle collision (*Run Over Ratio*: $M = 0.424$, $SD = 0.097$); this measure exhibited a relatively symmetric distribution.

Having examined the overall distributions of these behavioral measures, we next explored how the associations between self-reported risk-taking propensity and road-crossing

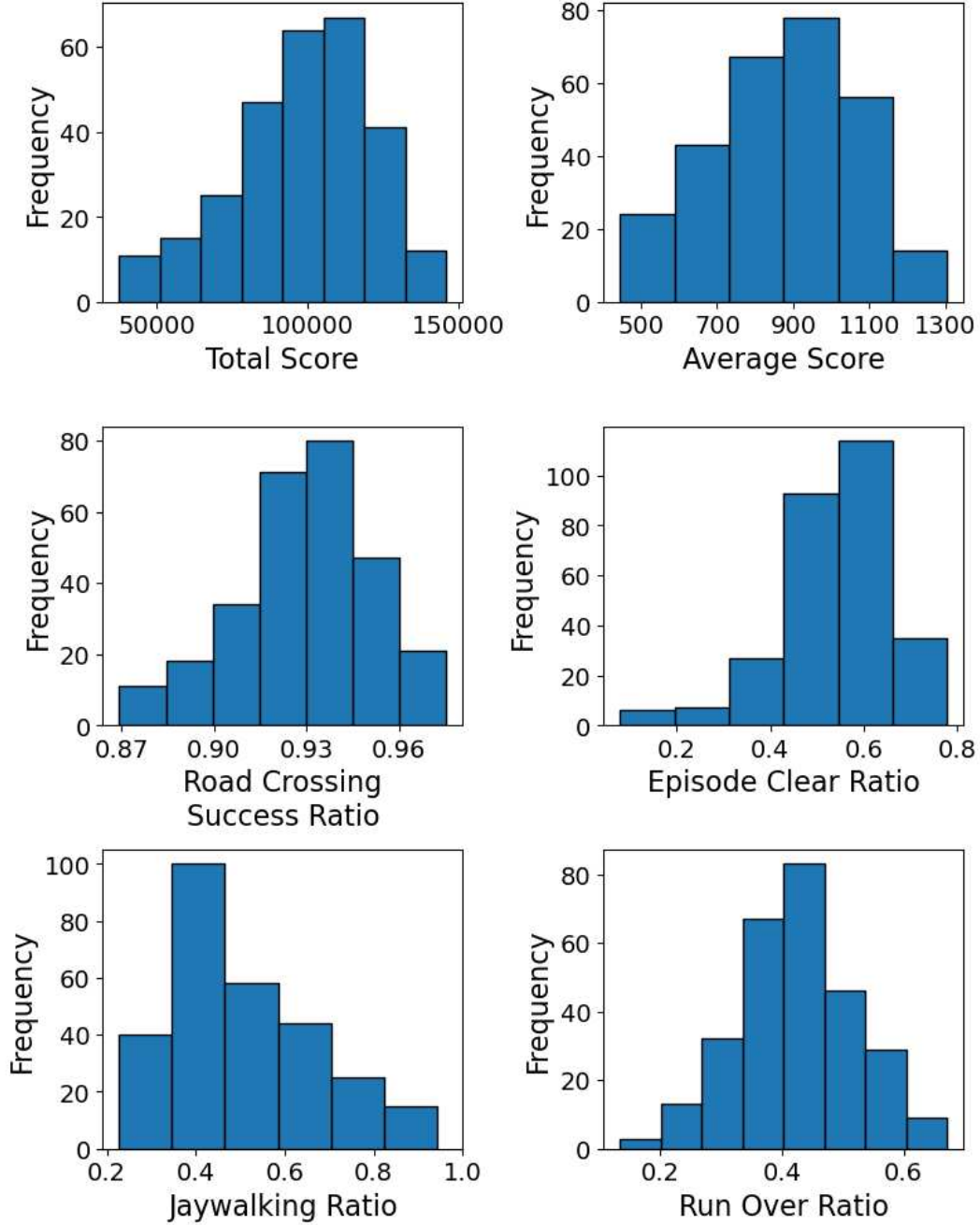


Figure 4. Distributions of six behavioral measures in the combined sample ($N = 282$).

behavior varied across collision penalty conditions. Specifically, we calculated Pearson correlations between DOSPERT scores and two behavioral measures, *Road Crossing Success Ratio* and *Run Over Ratio*, separately for each penalty level (Fig. 5).

In both the web sample and the combined sample, correlations with DOSPERT scores became progressively more negative for *Road Crossing Success Ratio* and more positive for

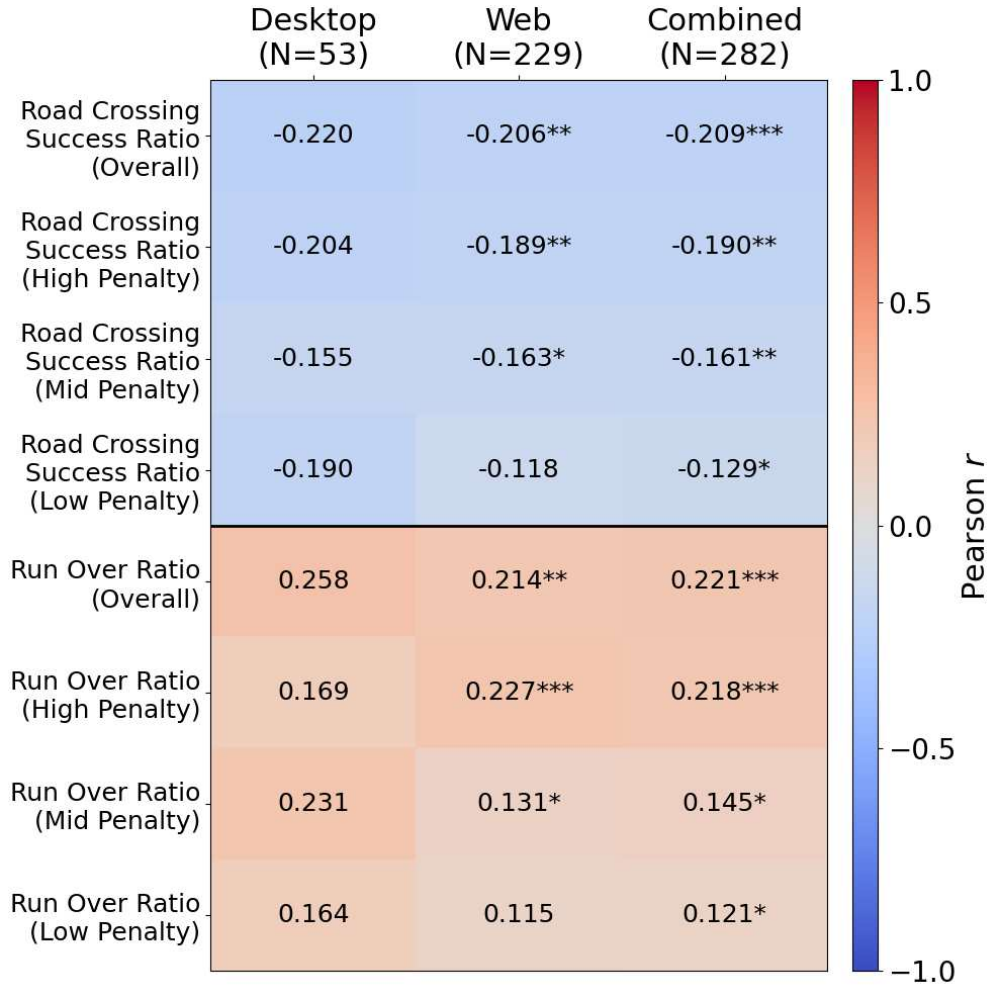


Figure 5. Pearson correlations of DOSPERT scores with *Road Crossing Success Ratio* and *Run Over Ratio* across different penalty conditions.

Run Over Ratio as collision penalties increased from low to medium to high. This pattern suggests that individual differences in risk-taking propensity can be more strongly reflected in observed road-crossing behavior under higher collision penalties.

IRL-Based Reward Inference

AIRL Model Training

AIRL was used to infer participant-specific reward functions from behavioral trajectories collected in the Crosswalk Task. Because the Crosswalk Task was implemented using the

Gymnasium framework (Towers et al., 2026), state–action trajectories were directly recorded at every time step, requiring no additional preprocessing. To augment the expert demonstrations, a left–right reflection of each trajectory was generated, yielding approximately 80 minutes of demonstration data per participant.

The AIRL algorithm was implemented using the imitation library (Gleave et al., 2022), with Proximal Policy Optimization (PPO; Schulman et al., 2017) as the generator policy. All models were trained using the same hyperparameters and training procedure with the random seed fixed at 42 for reproducibility. In each adversarial round, the generator interacted with the environment for 512 steps, followed by a single discriminator update using a demonstration batch size of 256. The discriminator was optimized using Adam with a learning rate of 1×10^{-3} and a weight decay of 5×10^{-4} . The reward network consisted of two hidden layers with 8 units each and LeakyReLU activations. The PPO policy comprised three hidden layers, each with 256 units and LeakyReLU activations, and was trained with a learning rate of 1×10^{-4} , a rollout length of 512, a minibatch size of 32, 8 optimization epochs per update, and a discount factor of 0.99. Observation normalization was not applied because all 113 state features generated by *CrosswalkEnv* were inherently bounded between 0 and 1; omitting normalization also preserved the original feature scales for subsequent interpretation of the inferred reward functions.

Among the saved checkpoints, the one with discriminator accuracy closest to chance-level performance (50% accuracy) was selected as the final model, indicating reduced distinguishability between expert demonstrations and generated trajectories. The selected checkpoints were typically obtained after 10,000–40,000 adversarial rounds. Using this procedure, participant-specific reward functions were inferred for 282 participants, whereas training runs that failed to reach the convergence criterion after an extended training period were considered training failures.

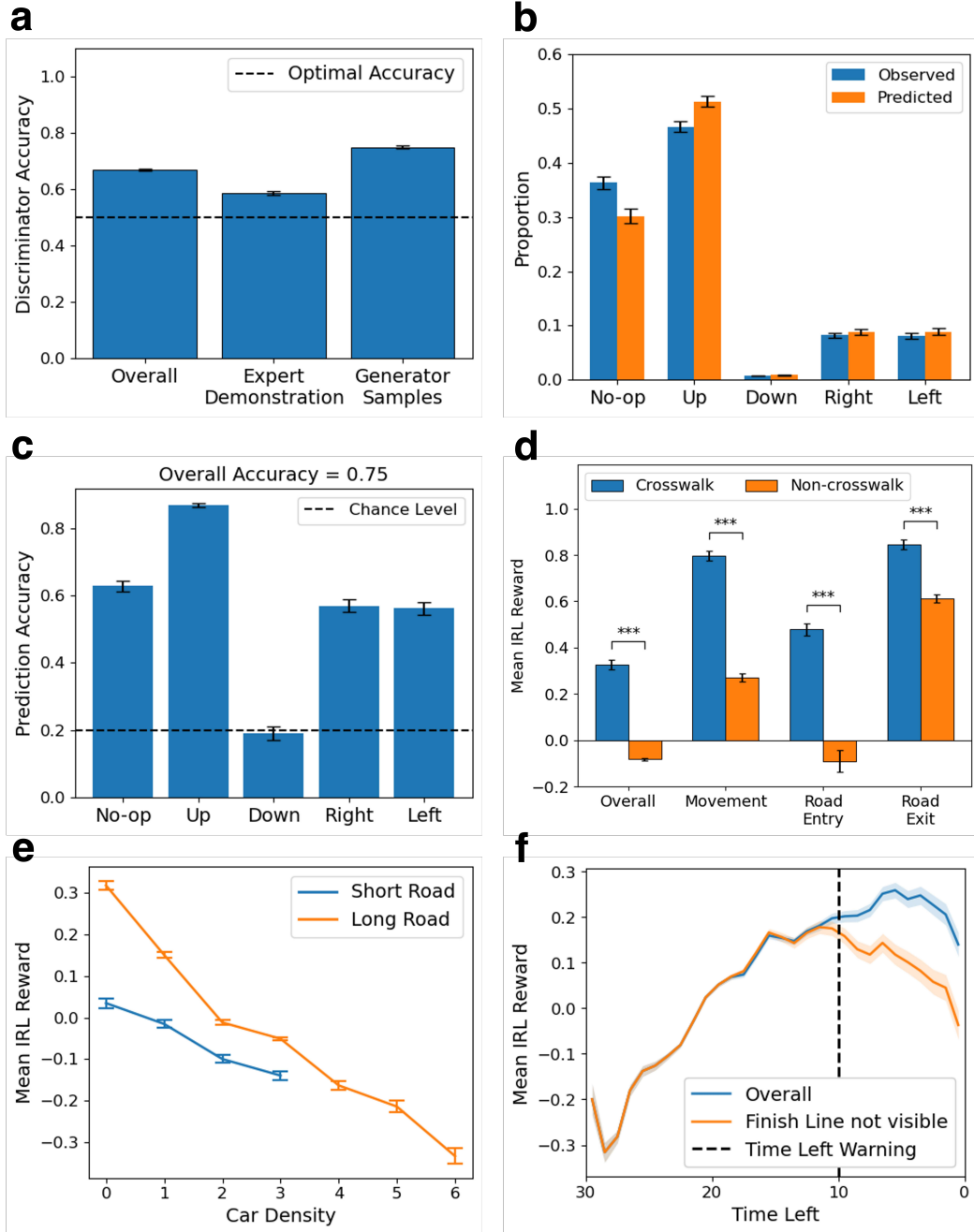


Figure 6. AIRL model evaluation. The plots show (a) discriminator accuracy, (b) observed and predicted action proportions, (c) action prediction accuracy, (d) mean rewards for crosswalk and non-crosswalk states, (e) mean rewards by car density and road length, and (f) mean rewards over time. Error bars and shaded areas indicate 95% confidence intervals. *** $p < .001$.

IRL Model Evaluation

Before interpreting the inferred reward functions, we evaluated the AIRL models in terms of discriminator performance, generator behavioral fit, and reward interpretability.

Discriminator In AIRL, an accuracy of 0.50 represents the theoretical equilibrium at which expert and generated trajectories are indistinguishable (Fu et al., 2018). The mean discriminator accuracy across the 282 trained models was 0.67. The discriminator accuracy on expert samples (0.59) was closer to the equilibrium than that on generated samples (0.75), suggesting that the generators captured some characteristics of expert behavior while producing trajectories with features that remained distinguishable from the expert trajectories (Fig. 6a). However, because discriminator accuracy alone is insufficient to establish the adequacy of model fit, additional analyses of behavioral fit and reward interpretability were conducted.

Generator The generators trained within the AIRL models were used to predict an action for each state in the participants’ observed trajectories.

The overall distribution of predicted actions closely matched the observed distribution across all five action types (Fig. 6b). In both distributions, upward movements were most frequent, followed by no-op actions, whereas rightward and leftward movements occurred at similar rates and downward movements were least frequent. This ordering reflected the task structure, as upward movements were required to reach the goal, whereas downward movements carried the player farther from it, resulting in an imbalanced action distribution. Modest action-level differences remained, primarily because participants produced more no-op responses than the AIRL-trained generators. Consistent with prior studies, this discrepancy may reflect human fatigue, inattention, and physical response constraints that are not fully captured by learned policies (Cross et al., 2021; Lee et al., 2024). Consequently, the higher observed rate of inaction reduced the proportions of the remaining actions relative

to their predicted proportions.

Beyond comparing the overall distributions of observed and predicted actions, we quantified how accurately the generators predicted the observed action for each state. The generators correctly predicted the observed actions with an overall accuracy of 75%, well above the chance level of 20% (Fig. 6c). The 95% confidence intervals for action-specific prediction accuracy, calculated as recall for each action, remained above chance for all action types except downward movement, for which the interval included the chance level. This lower performance was likely due to the rarity of downward actions in the observed trajectories, which resulted in an imbalanced dataset and a limited number of training examples for this action.

Reward Function Because the primary goal of AIRL was to recover behaviorally meaningful reward functions, we examined whether the inferred rewards exhibited plausible, task-relevant patterns.

The learned reward functions assigned consistently higher values to states and actions involving crosswalk use than to their non-crosswalk counterparts (Fig. 6d). Overall, a paired-samples t test showed that crosswalk states received significantly higher inferred rewards than non-crosswalk states ($t(281) = 32.59, p < .001$). This pattern was consistently observed for entering roads, moving around them, and exiting them ($ts(281) = 15.85\text{--}32.38, ps < .001$). Given that crosswalk use substantially reduces the risk of penalties relative to jaywalking, this pattern is consistent with the task structure and suggests that the inferred reward functions captured participants' preference for the safety afforded by crosswalk use.

The inferred rewards were also sensitive to the number of cars on the road being crossed. For both short and long road conditions, mean rewards were highest when there were no cars on the road but consistently decreased as the number of cars increased (Fig. 6e). Because a greater number of cars typically increases collision risk during road crossing, this pattern suggests that the inferred reward functions reflected participants' aversion to risky

situations.

Finally, the inferred rewards changed systematically with the time remaining in the episode (Fig. 6f). Inferred rewards generally increased as participants avoided collisions and progressed through the task, but declined when little time remained and the end of the map was still not visible. Crossing all roads on the map within the time limit yielded a bonus and maximized the total score, whereas remaining far from the goal near the deadline reduced the likelihood of completion. This pattern suggests that the inferred reward functions captured participants’ expected return under temporal constraints.

Risk Preference Indicators

Indicator Selection

To evaluate whether behavior in the Crosswalk Task reflects participants’ trait-level risk-taking propensity, we examined associations between task measures and participants’ DOSPERT total scores. Although the DOSPERT comprises five domain-specific subscales, the total score was used as the primary criterion for indicator selection because the analysis focused on general risk-taking propensity rather than domain-specific risk preferences.

Correlations were computed separately for the desktop, web, and combined samples. A task measure was identified as a risk preference indicator if it was significantly correlated with the DOSPERT total score in the combined sample and if the direction of the correlation was consistent across the desktop and web samples. Because the desktop sample was substantially smaller than the web sample, statistical significance in the desktop sample was interpreted with caution rather than treated as a strict selection criterion.

Applying these criteria, we identified five behavioral measures and six IRL reward measures as indicators of risk preference (Fig. 7; see the Supplementary Information for detailed definitions of the individual measures). Of these eleven measures, five were positively associated with self-reported risk-taking propensity (risk-taking indicators), whereas six were

negatively associated (risk-averse indicators).

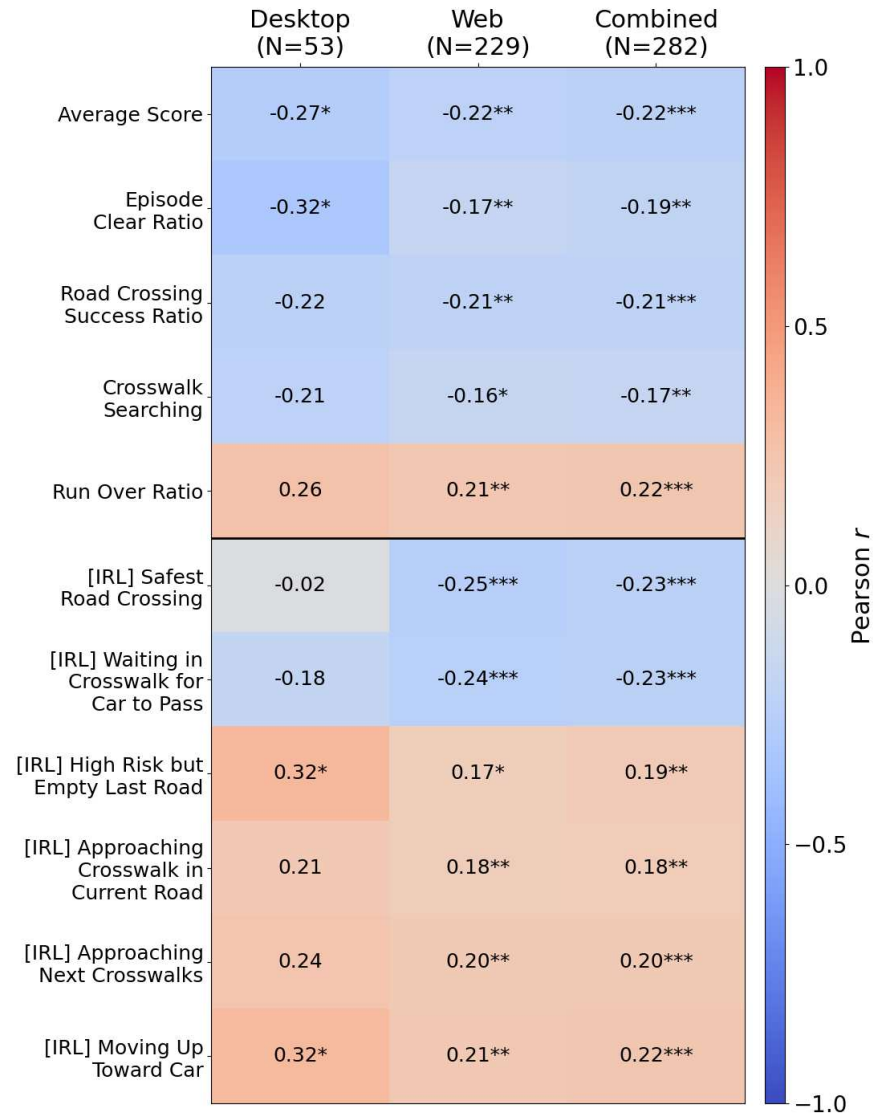


Figure 7. Correlations between DOSPERT total scores and the risk preference indicators from the Crosswalk Task in the desktop, web, and combined samples. Values represent Pearson's r . * $p < .05$, ** $p < .01$, *** $p < .001$.

Validation

Additional analyses were conducted to examine the construct validity of the selected indicators. Because consistency across the desktop and web implementations had already been assessed during indicator selection, these analyses were conducted using the combined sample

($N = 282$) to obtain more precise correlation estimates.

All selected indicators were significantly correlated with at least one DOSPERT domain score in the same direction as their correlations with the total score (Fig. 8). Most indicators were significantly associated with four or five domains, whereas one behavioral indicator was associated with only a single domain. None of the behavioral indicators was associated with the Social domain, whereas four of the six IRL reward indicators were associated with all five domains. Taken together, these findings suggest that the selected indicators, particularly the IRL reward indicators, reflect a relatively general propensity for risk taking rather than preferences confined to a specific domain.



Figure 8. Correlations between the selected risk preference indicators and the DOSPERT total and domain scores in the combined sample ($N = 282$). Values represent Pearson's r . * $p < .05$, ** $p < .01$, *** $p < .001$.

To evaluate discriminant validity, we examined whether the selected risk preference indicators were associated specifically with self-reported risk taking rather than with related psychological constructs. Prior research has shown that risk taking may be associated with impulsivity (Lauriola et al., 2014; Nigg, 2017) and anxiety (Price et al., 2025; Wake et al., 2020). Consistent with this literature, DOSPERT total scores in the combined sample were positively correlated with both BIS-11 scores ($r = .23, p < .001$) and STAI scores ($r = .13, p < .05$), indicating modest overlap among self-reported risk taking, impulsivity, and anxiety. Despite this overlap at the questionnaire level, the selected risk preference indicators generally showed little systematic association with BIS-11 or STAI scores (see Fig. S4 in the Supplementary Information). None of the indicators was significantly correlated with either state or trait anxiety, and no significant associations were observed with the BIS-11 total or subscale scores, except for two indicators that were correlated with the BIS-11 Motor subscale. Overall, this pattern suggests that the selected task indicators were more specifically related to risk-taking propensity than to impulsivity or anxiety, providing evidence for their discriminant validity.

Multivariate Results

The preceding analyses identified several task measures that were individually associated with self-reported risk-taking propensity. We next examined whether jointly considering these indicators could provide information about individual differences in risk taking beyond that provided by their separate associations with self-reported risk-taking propensity. To address this question, we used multiple machine learning approaches to jointly analyze the selected risk preference indicators.

Regression Analysis

To evaluate the predictive information captured by the selected indicators, we used elastic net regression (Zou and Hastie, 2005) to predict DOSPERT scores from measures derived from the Crosswalk Task. The mixing parameter was set to $\alpha = 0.5$ to balance feature selection and coefficient stability by assigning equal weight to L1 and L2 regularization. To compare the informativeness of different types of risk preference indicators, three models were trained and compared: a behavioral feature model, an IRL feature model, and a combined model that included all 11 measures. Models were trained on 80% of the data using five-fold cross-validation and tested on the remaining 20%. Predictive performance was quantified as the correlation between predicted and observed DOSPERT scores in the test set.



Figure 9. Beta coefficient estimates for elastic net models predicting DOSPERT scores using different sets of predictors. (a) The behavioral model included five behavioral measures. (b) The IRL model included six IRL reward measures. (c) The combined model included all 11 measures: five behavioral measures and six IRL reward measures. Points indicate the mean coefficient estimates, and error bars represent intervals defined as the mean ± 1.96 SD. Predictors whose coefficients were shrunk to zero by the elastic net were not retained in the corresponding model.

Figure 9 summarizes the feature selection results across the elastic net models. The behavioral-only model retained three of the five behavioral measures and showed a weak correlation with DOSPERT scores (Fig. 9a; $r = .23$), providing no evidence of improvement beyond the correlations observed for individual behavioral measures ($|r| = .14-.22$; see Fig. 7). In contrast, the IRL-only model retained all six IRL reward measures and achieved a higher correlation (Fig. 9b; $r = .31$), exceeding the correlations observed for individual IRL reward measures ($|r| = .18-.23$). The combined model achieved the highest correlation (Fig. 9c; $r = .36$), retaining all six IRL reward measures but only two of the five behavioral measures. Together, these patterns suggest that predictive information is distributed across multiple IRL reward measures and that these measures provide greater predictive information when considered jointly, whereas combining behavioral measures provides little gain beyond their individual associations with DOSPERT scores.

To examine whether the observed predictive performance generalized beyond a specific algorithm and its linearity assumption, we also trained a multilayer perceptron (MLP) to predict DOSPERT scores using all 11 measures. Across five repetitions of five-fold cross-validation, the MLP showed moderate predictive performance ($r = .3121 \pm .1206$), comparable to that of the combined elastic net model ($r = .33$). Together, these results suggest that the Crosswalk Task measures provide predictive information across both linear and nonlinear modeling approaches. Detailed MLP configuration and results are reported in Table S4 in the Supplementary Information.

Clustering Analysis

To assess whether behavior in the Crosswalk Task could distinguish participant groups with different risk preferences without using self-report measures, we applied unsupervised k-means clustering (McQueen, 1967) to the combined sample using only the eleven risk preference indicators derived from the task. We selected two clusters based on the highest mean silhouette score (Fig. S7a) and an elbow in the inertia curve near $k = 2$ (Fig. S7b; see the

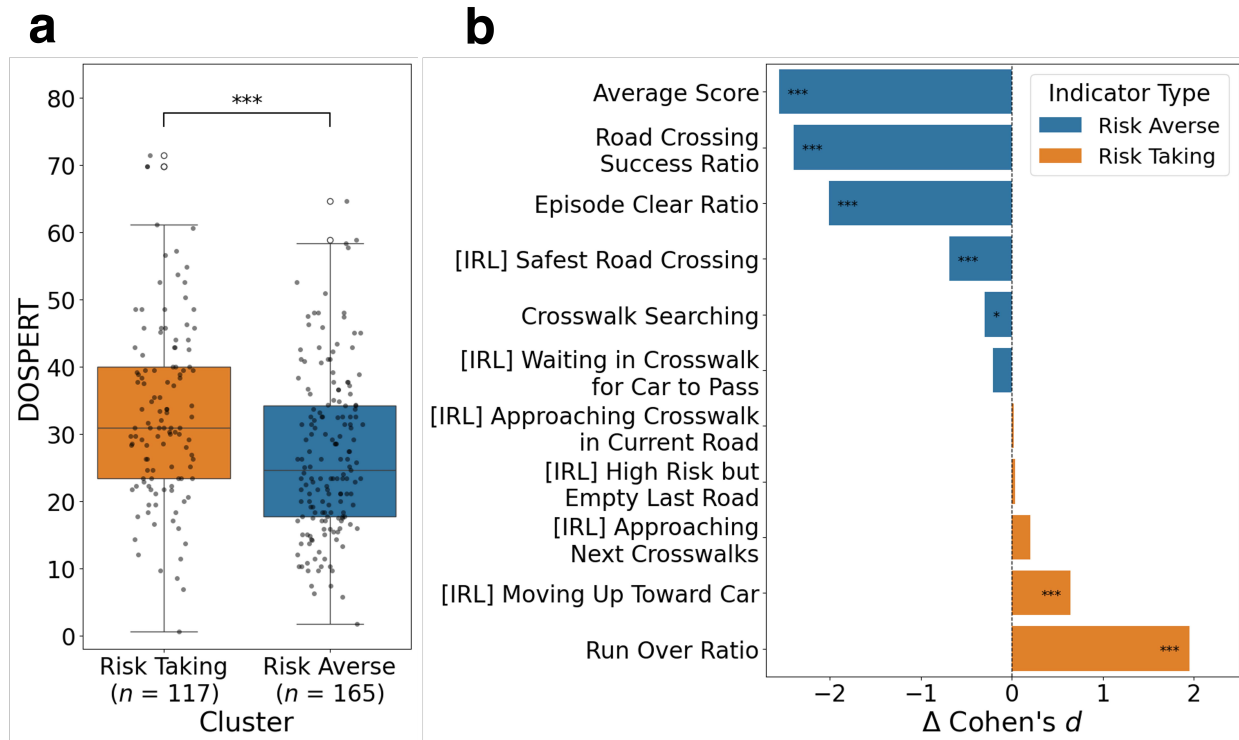


Figure 10. Characterization and external validation of the identified behavioral clusters. (a) DOSPERT score distributions across clusters, which were labeled Risk Averse and Risk Taking based on their relative DOSPERT scores. (b) Standardized differences in risk preference indicators, calculated as Risk Taking minus Risk Averse, such that positive values indicate higher indicator values in the Risk Taking cluster. * $p < .05$, *** $p < .001$.

Supplementary Information for details).

DOSPERT scores, which were not used to determine either the number of clusters or cluster assignments, differed significantly between the resulting groups (Welch's $t(232.5) = 4.12$, $p < .001$, $d = 0.51$; Fig. 10a). We therefore labeled the groups *Risk Taking* and *Risk Averse* based on their relative DOSPERT scores. This correspondence with an external self-report measure supports the validity of the behaviorally derived clusters and suggests that the Crosswalk Task can distinguish groups with different risk preferences using task-derived indicators alone.

The clusters also exhibited consistent behavioral profiles across the eleven input indicators. All indicators showed differences in the expected direction, with the *Risk Taking* cluster exhibiting higher values on risk-taking indicators and the *Risk Averse* cluster ex-

hibiting higher values on risk-averse indicators; seven of these differences were statistically significant (Fig. 10b).

Discussion

The present study introduces the Crosswalk Task as a naturalistic reinforcement learning paradigm for measuring individual differences in risk taking. Across the desktop sample assessed in the laboratory and the web sample assessed online, multiple behavioral and IRL-derived measures were associated with self-reported risk taking in consistent directions, yielding eleven indicators of risk preference. These indicators were associated with multiple domains of the DOSPERT while showing little systematic association with related psychological constructs, providing evidence for their construct and discriminant validity.

Moreover, considering multiple risk preference indicators together provided meaningful information about individual differences in risk taking beyond that provided by individual measures. Predictions from multivariate models showed moderate correspondence with DOSPERT scores, and similar performance across linear and nonlinear approaches suggested that this pattern was not specific to linear models. Complementing these predictive results, unsupervised clustering based solely on the task measures identified groups that differed significantly in DOSPERT scores, even though DOSPERT scores were not used to define the clusters. The *Risk Taking* cluster also showed consistently higher values on risk-taking indicators and lower values on risk-averse indicators than the *Risk Averse* cluster, suggesting that these indicators provide a useful feature space for unsupervised stratification of individual risk preference profiles. Together, these findings suggest that integrating multiple task measures can provide a more informative characterization of risk preference than considering each measure in isolation.

Notably, IRL reward measures consistently provided a richer characterization of risk taking than simple aggregate behavioral measures. They showed associations with a broader

range of DOSPERT subscales (Fig. 8). In the elastic net analysis, all six IRL reward measures were retained, suggesting that each provided unique predictive information and that, jointly, they contributed to higher predictive performance than behavioral measures alone (Fig. 9). These findings support prior work suggesting that deep IRL can provide a practical framework for leveraging complex real-time behavioral data beyond simple descriptive summaries (Lee et al., 2024).

Methodologically, the Crosswalk Task extends prior naturalistic real-time paradigms by introducing *CrosswalkEnv*, a new Gymnasium environment developed specifically for behavioral research. Unlike previous studies that modified existing programs to create behavioral tasks, this open-source implementation allows researchers to extend the environment and flexibly design their own tasks. Importantly, implementing the task within the Gymnasium framework allows behavior to be represented naturally as sequential state-action trajectories, enabling RL-based analyses of human behavior.

Accordingly, the Crosswalk Task may be particularly useful when researchers seek to characterize individual differences using multiple aspects of real-time behavior rather than relying on a single summary measure. Beyond risk preference, modified versions of the task could be used to investigate other psychological constructs. The paradigm may also be well suited for neuroimaging studies, as prior work has demonstrated the feasibility of combining real-time behavioral tasks with IRL-derived rewards to investigate the neural correlates of these rewards (Lee et al., 2026).

We emphasize that the eleven risk preference indicators examined in this study represent only a subset of the measures that could be derived from the task. By making the software publicly available, we aim to encourage researchers to explore additional aspects of behavior and develop new measures.

Nevertheless, several limitations regarding data collection, analysis, and validation should be considered. First, comparisons with other measures of risk taking were limited in scope. Exploratory comparisons with two traditional behavioral tasks related to risk

taking showed that multiple Crosswalk Task measures were significantly associated with DOSPERT total scores, whereas the measures derived from the other two tasks were not. However, these comparisons were limited to the relatively small desktop sample ($N = 53$) and should therefore be interpreted with caution (see Fig. S5 in the Supplementary Information). Moreover, because the DOSPERT was the only self-report measure of trait-level risk-taking propensity, it remains unclear whether the identified indicators generalize beyond this specific measure. Future work should examine how these indicators relate to other established self-report measures, such as the one-item General Risk Taking Question (Dohmen et al., 2011) and the Risk Propensity Scale (Nicholson et al., 2005), and behavioral paradigms, such as the Balloon Analogue Risk Task (BART; Lejuez et al., 2002) and Iowa Gambling Task (IGT; Bechara et al., 1994).

Second, the desktop and web samples differed in sample size, language, nationality, and study setting. Although the observed associations were broadly consistent, this substantial heterogeneity limits the interpretation of the combined sample results. Furthermore, the combined sample results were driven primarily by the larger web sample ($N = 229$), with the small desktop sample ($N = 53$) contributing relatively little. Future work should verify whether these findings are consistent across online and laboratory settings within the same population. Replication in larger and more diverse laboratory samples is particularly important for future neuroimaging studies requiring specialized hardware, such as MRI scanners.

Finally, participants were not clinically characterized, despite substantial evidence of differences in decision-making processes between psychiatric and healthy populations (Amlung et al., 2019; Pike and Robinson, 2022). Future research should evaluate the Crosswalk Task across various clinical groups and healthy controls to assess its clinical validity, thereby contributing to efforts to develop behavioral tasks with greater clinical utility in mental health research (Ahn and Busemeyer, 2016).

Taken together, the present findings indicate that the Crosswalk Task provides a naturalistic reinforcement learning framework for characterizing individual differences in risk

taking from real-time behavior. Its Gymnasium-based implementation produces structured state-action trajectories that enable RL-based analyses and supports flexible task modification, while its complementary web-based implementation supports scalable data collection. Diverse task measures can be derived from these rich behavioral trajectories, and computational approaches such as IRL can further quantify how individuals value specific decision contexts. The Crosswalk Task thus lays the foundation for developing flexible, naturalistic reinforcement learning paradigms for studying decision making.

Declarations

Funding

This work was supported by the National Research Foundation of Korea (NRF) (RS-2024-00435727, RS-2025-00516410, and RS-2022-KH125035), the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (RS-2026-25507282), the Seoul National University Artificial Intelligence Graduate School Program (RS-2021-II211343), K-NeuroMind NEXUS: Neurocognitive-Affective Next-Generation Cross-modal Foundation Platform (RS-2026-25518987), and the BK21 FOUR Program (5199990314123) to WYA. The funders had no role in study design, data collection, data analysis, interpretation of the data, or preparation of the manuscript.

Competing Interests

The authors declare no competing interests.

Ethics approval

The data collection procedures for the in-person and online experiments were approved by the Institutional Review Board of Seoul National University (IRB Nos. 2509/002-005

and 2603/003-037, respectively). All procedures were conducted in accordance with the Declaration of Helsinki and its later amendments.

Consent to participate

Written informed consent was obtained from all participants in the in-person experiment. For the online experiment, informed consent was obtained electronically through an online consent checkbox. Both consent procedures were approved by the Institutional Review Board (IRB) prior to data collection.

Consent for publication

Written consent for the publication of anonymized data was obtained from all participants in the in-person experiment. For the online experiment, consent for publication was obtained electronically through an online consent checkbox.

Availability of data and materials

The raw data collected from 309 participants and the analysis code are publicly available in the GitHub repository at <https://github.com/CCS-Lab/crosswalk-task-public>.

The Supplementary Information is available on the Open Science Framework (OSF) at <https://osf.io/ap4z7/>.

Code availability

Both software packages developed in this study are publicly available on GitHub. The Python package *CrosswalkEnv* is available at <https://github.com/crosswalk-task/CrosswalkEnv>, and the npm package *CrosswalkWebEnv* is available at <https://github.com/crosswalk-task/CrosswalkWebEnv>.

Detailed documentation for both packages is available on the project website at <https://>

[//crosswalk-task.github.io/](https://crosswalk-task.github.io/).

Open Practices Statement

Materials and analysis code are publicly available on GitHub at <https://github.com/CS-Lab/crosswalk-task-public>. The Supplementary Information is available on OSF at <https://osf.io/ap4z7/>. None of the reported studies were preregistered.

References

- Aghabayk, K., Esmailpour, J., Jafari, A., & Shiwakoti, N. (2021). Observational-based study to explore pedestrian crossing behaviors at signalized and unsignalized crosswalks. *Accident Analysis & Prevention*, 151, 105990.
- Ahn, W.-Y., & Busemeyer, J. R. (2016). Challenges and promises for translating computational tools into clinical practice. *Current Opinion in Behavioral Sciences*, 11, 1–7.
- Allen, K., Brändle, F., Botvinick, M., Fan, J. E., Gershman, S. J., Gopnik, A., Griffiths, T. L., Hartshorne, J. K., Hauser, T. U., Ho, M. K., et al. (2024). Using games to understand the mind. *Nature human behaviour*, 8(6), 1035–1043.
- Amlung, M., Marsden, E., Holshausen, K., Morris, V., Patel, H., Vedelago, L., Naish, K. R., Reed, D. D., & McCabe, R. E. (2019). Delay discounting as a transdiagnostic process in psychiatric disorders: A meta-analysis. *JAMA psychiatry*, 76(11), 1176–1186.
- Arora, S., & Doshi, P. (2021). A survey of inverse reinforcement learning: Challenges, methods and progress. *Artificial Intelligence*, 297, 103500.
- Bagaïni, A., Liu, Y., Kapoor, M., Son, G., Bürkner, P.-C., Tisdall, L., & Mata, R. (2025). A systematic review and meta-analyses of the temporal stability and convergent validity of risk preference measures. *Nature Human Behaviour*, 9(4), 700–712.
- Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50(1-3), 7–15.
- Blais, A.-R., & Weber, E. U. (2006). A domain-specific risk-taking (dospert) scale for adult-populations. *Judgment and Decision making*, 1(1), 33–47.
- Bran, A., & Vaidis, D. C. (2020). Assessing risk-taking: What to measure and how to measure it. *Journal of Risk Research*, 23(4), 490–503.

- Cœugnet, S., Cahour, B., & Kraiem, S. (2019). Risk-taking, emotions and socio-cognitive dynamics of pedestrian street-crossing decision-making in the city. *Transportation research part F: traffic psychology and behaviour*, 65, 141–157.
- Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for pomp. *Multivariate behavioral research*, 34(3), 315–346.
- Cross, L., Cockburn, J., Yue, Y., & O’Doherty, J. P. (2021). Using deep reinforcement learning to reveal how the brain encodes abstract state-space representations in high-dimensional environments. *Neuron*, 109(4), 724–738.
- Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated? *Trends in cognitive sciences*, 24(4), 267–269.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association*, 9(3), 522–550.
- Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the columbia card task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 709.
- Figner, B., & Weber, E. U. (2011). Who takes risks when and why? determinants of risk taking. *Current directions in psychological science*, 20(4), 211–216.
- Fox, C. R., & Tannenbaum, D. (2011). The elusive search for stable risk preferences. *Frontiers in psychology*, 2, 298.
- Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science advances*, 3(10), e1701381.
- Fu, J., Luo, K., & Levine, S. (2018). Learning robust rewards with adversarial inverse reinforcement learning. *International Conference on Learning Representations (ICLR)*.

- Gleave, A., Taufeeque, M., Rocamonde, J., Jenner, E., Wang, S. H., Toyer, S., Ernestus, M., Belrose, N., Emmons, S., & Russell, S. (2022). Imitation: Clean imitation learning implementations. *arXiv preprint arXiv:2211.11972*.
- Hahn, D.-W., Lee, C.-H., & Chon, K.-K. (1996). Korean adaptation of spielberger's stai (k-stai). *Korean journal of health psychology*, 1(1), 1–14.
- Hashemiparast, M., Sharma, M., Asghari Jafarabadi, M., & Hosseini, Z. (2020). Still careless: Findings from a cross-sectional study of young pedestrians' risky road crossing behaviors. *Archives of public health*, 78(1), 44.
- Hatfield, J., Williamson, A., Kehoe, E. J., & Prabhakharan, P. (2017). An examination of the relationship between measures of impulsivity and risky simulated driving amongst young drivers. *Accident Analysis & Prevention*, 103, 37–43.
- Heo, S. Y., Kim, J. H., & Oh, J. Y. (2012). The korean version of the barratt impulsiveness scale, 11th version: Its reliability and validity. *Korean Journal of Psychology: General*, 31(3), 769–782.
- Hong, J. I., Lee, Y., Kim, K., Park, J., Park, S., Lee, K., & Lee, J. (2010). Risk-taking propensity of korean people that were measured by bart and k-dospert. *Journal of Korean Association of Social Psychiatry*, 14, 3–14.
- Lauriola, M., Panno, A., Levin, I. P., & Lejuez, C. W. (2014). Individual differences in risky decision making: A meta-analysis of sensation seeking and impulsivity with the balloon analogue risk task. *Journal of Behavioral Decision Making*, 27(1), 20–36.
- Lee, S. H., Chung, C., Oh, M.-h., & Ahn, W.-Y. (2026). Inverse reinforcement learning reveals action-oriented value signals in naturalistic decision making. *bioRxiv*, 2026–06.
- Lee, S. H., Song, M. S., Oh, M.-h., & Ahn, W.-Y. (2024). Bridging the gap between self-report and behavioral laboratory measures: A real-time driving task with inverse reinforcement learning. *Psychological Science*, 35(4), 345–357.
- 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.
- Liu, Y.-C., & Tung, Y.-C. (2014). Risk analysis of pedestrians' road-crossing decisions: Effects of age, time gap, time of day, and vehicle speed. *Safety science*, 63, 77–82.
- McQueen, J. B. (1967). Some methods of classification and analysis of multivariate observations. *Proc. of 5th berkeley symposium on math. stat. and prob.*, 281–297.
- Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: Rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, 222, 117254.
- Ng, A. Y., Russell, S., et al. (2000). Algorithms for inverse reinforcement learning. *Icml*, 1(2), 2.
- Nicholson, N., Soane, E., Fenton-O'Creevy, M., & Willman, P. (2005). Personality and domain-specific risk taking. *Journal of Risk Research*, 8(2), 157–176.
- Nigg, J. T. (2017). Annual research review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of child psychology and psychiatry*, 58(4), 361–383.
- Niv, Y. (2009). Reinforcement learning in the brain. *Journal of mathematical psychology*, 53(3), 139–154.
- Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. *Journal of Clinical Psychology*, 51(6), 768–774.
- Pike, A. C., & Robinson, O. J. (2022). Reinforcement learning in patients with mood and anxiety disorders vs control individuals: A systematic review and meta-analysis. *JAMA psychiatry*, 79(4), 313–322.
- Piyalungka, S., Kanitpong, K., & Karoonsoontawong, A. (2025). Pedestrian gap acceptance behavior at unsignalized mid-block crossing under mixed traffic conditions. *IATSS Research*, 49(2), 105–113.

- Price, P. C., de Jesus, Y., Fernandez, L., & Hurd, J. (2025). Do anxious people take fewer risks? a meta-analysis of the relationship between trait anxiety and dispositional risk taking. *Journal of Behavioral Decision Making*, 38(5), e70054.
- Schonberg, T., Fox, C. R., & Poldrack, R. A. (2011). Mind the gap: Bridging economic and naturalistic risk-taking with cognitive neuroscience. *Trends in cognitive sciences*, 15(1), 11–19.
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.
- Slovic, P. (1962). Convergent validation of risk taking measures. *The Journal of Abnormal and Social Psychology*, 65(1), 68.
- Slovic, P. (1964). Assessment of risk taking behavior. *Psychological bulletin*, 61(3), 220.
- Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory (form y)*. Consulting Psychologists Press.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.
- Theofilatos, A., Ziakopoulos, A., Oviedo-Trespalacios, O., & Timmis, A. (2021). To cross or not to cross? review and meta-analysis of pedestrian gap acceptance decisions at midblock street crossings. *Journal of Transport & Health*, 22, 101108.
- Tian, K., Markkula, G., Wei, C., Lee, Y. M., Madigan, R., Hirose, T., Merat, N., & Romano, R. (2023). Deconstructing pedestrian crossing decisions in interactions with continuous traffic: An anthropomorphic model. *IEEE Transactions on Intelligent Transportation Systems*, 25(3), 2466–2478.
- Towers, M., Kwiatkowski, A., Balis, J., De Cola, G., Deleu, T., Goulão, M., Andreas, K., Krimmel, M., Kg, A., Perez-Vicente, R., et al. (2026). Gymnasium: A standard interface for reinforcement learning environments. *Advances in Neural Information Processing Systems*, 38.
- Wake, S., Wormwood, J., & Satpute, A. B. (2020). The influence of fear on risk taking: A meta-analysis. *Cognition and Emotion*, 34(6), 1143–1159.

- Wang, Y., Niu, Y., Zhu, W., Chen, W., Li, Q., & Wang, T. (2023). Predicting pedestrian crossing behavior at unsignalized mid-block crosswalks using maximum entropy deep inverse reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*, 25(5), 3685–3698.
- Wang, Y., Srinivasan, A. R., Jokinen, J. P., Oulasvirta, A., & Markkula, G. (2023). Modeling human road crossing decisions as reward maximization with visual perception limitations. *2023 IEEE Intelligent Vehicles Symposium (IV)*, 1–6.
- Wang, Y., Srinivasan, A. R., Lee, Y. M., & Markkula, G. (2025). Modeling pedestrian crossing behavior: A reinforcement learning approach with sensory motor constraints. *IEEE Transactions on Intelligent Transportation Systems*, 26(10), 16454–16465.
- Wise, T., Emery, K., & Radulescu, A. (2024). Naturalistic reinforcement learning. *Trends in Cognitive Sciences*, 28(2), 144–158.
- Wulfmeier, M., Ondruska, P., & Posner, I. (2015). Maximum entropy deep inverse reinforcement learning. *arXiv preprint arXiv:1507.04888*.
- Yannis, G., Papadimitriou, E., & Theofilatos, A. (2013). Pedestrian gap acceptance for mid-block street crossing. *Transportation planning and technology*, 36(5), 450–462.
- Ziebart, B. D., Maas, A. L., Bagnell, J. A., Dey, A. K., et al. (2008). Maximum entropy inverse reinforcement learning. *Aaai*, 8, 1433–1438.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2), 301–320.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [CrosswalkTaskSupplementary.pdf](#)