

Supplementary Information

S1 Gymnasium Environment

The Crosswalk Task uses an internal Gymnasium environment with a discrete action space comprising five actions: up, down, left, right, and no-op. Every action advances the environment by one time step. With the no-op action, the agent stays in place while the cars continue to move. At each time step t , the agent selects an action a_t and receives a 113-dimensional observation vector $\mathbf{o}_t$:

$$a_t \in \mathcal{A} = \{a_{\text{no-op}}, a_{\text{up}}, a_{\text{down}}, a_{\text{right}}, a_{\text{left}}\},$$

$$\mathbf{o}_t \in \mathcal{O} = [0, 1]^{113}.$$

The observation vector encodes temporal and spatial context, together with road, crosswalk, and traffic information relative to the agent’s current position. All elements lie in $[0, 1]$, so no additional feature-wise normalization is required for reinforcement learning (RL; Sutton and Barto, 1998) or inverse reinforcement learning (IRL; Ng, Russell, et al., 2000) training. The observation vector comprises 82 continuous features, 17 binary features, and 14 discrete features. Table S1 groups the 113 features into five observational categories.

The observation vector is designed to capture information relevant to the agent’s current and near-future movement decisions. For example, the crosswalk-location features encode the relative positions of up to three visible crosswalks on the screen. These features allow the model to capture cases in which the agent moves laterally toward a crosswalk it plans to use later, while still crossing the current road. Likewise, the car observation features capture positional information about cars within the agent’s visible screen area across six rows: four ahead of the agent, its current row, and one behind it.

A complete index-by-index specification of the observation vector is available in the online documentation (https://crosswalk-task.github.io/crosswalkenv/obs_space.html).

Table S1. Composition of the 113-dimensional observation vector.

Indices	Category	Description
0–4	Temporal and spatial context	Normalized remaining episode time and the agent’s directional proximity to the map boundaries.
5–12	Nearby grid features	Binary indicators of whether cars may appear in the current, next, and previous rows and whether the agent’s current tile and the four adjacent tiles are crosswalk activation tiles.
13–23	Target-road information	Crosswalk visibility, collision penalty, distances from the start and end of the road, visible-car density, and directional counts of approaching cars on the target road.
24–32	Crosswalk locations	Normalized proximity to up to three visible crosswalks in the upward, rightward, and leftward directions, ordered from nearest to farthest.
33–112	Car observations	Eight binary indicators of car occupancy in the tiles surrounding the agent, and 12 position- and speed-related features for approaching cars in each of six rows relative to the agent.

S2 Inverse Reinforcement Learning Algorithm

Adversarial inverse reinforcement learning (AIRL; Fu et al., 2018) was used to infer participant-specific reward functions. AIRL implements IRL by alternating between discriminator training and policy optimization in an adversarial framework.

The goal of the discriminator is to distinguish transitions in observed trajectories (i.e., behavioral data collected from participants) from those generated by the current policy. The discriminator is defined as

$$D_{\theta,\phi}(s, a, s') = \frac{\exp(f_{\theta,\phi}(s, a, s'))}{\exp(f_{\theta,\phi}(s, a, s')) + \pi(a \mid s)}, \quad (1)$$

where s , a , s' , and $\pi(a \mid s)$ denote the current state, action, next state, and the probability of selecting action a under the current policy, respectively.

The function $f_{\theta,\phi}(s, a, s')$ is defined as

$$f_{\theta,\phi}(s, a, s') = g_{\theta}(s, a) + \gamma h_{\phi}(s') - h_{\phi}(s), \quad (2)$$

where g_θ is a reward function approximator, γ is the discount factor, and h_ϕ is a potential function used for reward shaping. The reward approximator $g_\theta(s, a)$ can be parameterized as a function of the state alone, $g_\theta(s)$, to encourage the recovery of reward functions that are less dependent on the specific dynamics of the task environment.

The policy is trained to maximize the estimated reward, which is derived from the discriminator as

$$\begin{aligned}\hat{r}(s, a, s') &= \log D_{\theta, \phi}(s, a, s') - \log (1 - D_{\theta, \phi}(s, a, s')) \\ &= f_{\theta, \phi}(s, a, s') - \log \pi(a | s).\end{aligned}\tag{3}$$

Through alternating discriminator and policy updates, the policy learns to generate transitions that resemble those observed in the participant data. As the policy-generated transition distribution approaches the observed distribution, the discriminator becomes increasingly unable to distinguish between the two. At the ideal adversarial equilibrium, the optimal discriminator outputs a probability of 0.5, corresponding to chance-level classification accuracy (50%) when observed and generated transitions are equally represented.

The AIRL training procedure is summarized in Algorithm 1. The procedure was applied separately to each participant to infer participant-specific reward functions.

Algorithm 1 Adversarial inverse reinforcement learning

- 1: Obtain observed trajectories τ_i^O for each participant.
 - 2: Initialize policy π and discriminator $D_{\theta, \phi}$.
 - 3: **for** step $t \in \{1, \dots, N\}$ **do**
 - 4: Sample trajectories $\tau_i = (s_0, a_0, \dots, s_T, a_T)$ using the policy π .
 - 5: Train $D_{\theta, \phi}$ via binary logistic regression to distinguish transitions in τ_i^O from those in τ_i .
 - 6: Update reward: $r_{\theta, \phi}(s, a, s') \leftarrow \log (D_{\theta, \phi}(s, a, s')) - \log (1 - D_{\theta, \phi}(s, a, s'))$.
 - 7: Update policy π with respect to $r_{\theta, \phi}$ using any policy optimization method.
 - 8: **end for**
-

S3 Data Collection

S3.1 Questionnaires

Participants completed five questionnaires: the Domain-Specific Risk-Taking Scale (DOSPERT; Blais and Weber, 2006), the Barratt Impulsiveness Scale (BIS-11; Patton et al., 1995), the State-Trait Anxiety Inventory (STAI; Spielberger et al., 1983), the Internet Gaming Disorder Scale (IGDS; Lemmens et al., 2015), and the Internet Addiction Test (IAT; Young, 2016). Scores on the IGDS and IAT were collected for potential use as covariates to account for associations between problematic gaming or internet use and behavioral-task performance, but neither measure was included in the final analyses. The DOSPERT, BIS-11, and STAI are described in the main text, whereas details of the IGDS and IAT are provided below.

Internet Gaming Disorder Scale Participants completed the 27-item IGDS to assess symptoms of internet gaming disorder. Responses were recorded on a six-point scale ranging from 1 (*never*) to 6 (*every day or almost every day*). Korean participants completed the K-IGDS (Cho and Kwon, 2017), a validated Korean version of the IGDS.

Internet Addiction Test Participants completed the 20-item IAT to assess the severity of problematic internet use. Items were rated on a five-point Likert-type scale ranging from 1 (*Rarely*) to 5 (*Always*). For the web sample, a *Not Applicable* option was added, and responses to this option were coded as 1 (*Rarely*) when calculating total scores; the possible total-score range was 20–100. Korean participants completed the K-IAT (Gyeong et al., 2012), a validated Korean version of the IAT.

S3.2 Data Screening and Participant Exclusion

The screening procedures were designed to identify participants whose data might be unreliable because of insufficient task engagement or inattentive responding. Because the desktop

experiment was conducted under direct experimenter supervision, data from the desktop sample were considered more reliable. In contrast, the web sample underwent additional screening procedures to address data-quality concerns particularly relevant in online settings without direct supervision (Rodd, 2024).

First, participants were screened based on their Crosswalk Task performance. As an initial within-sample check, each participant's best and average scores were standardized separately within the desktop and web samples. For the desktop sample, a z -score below -3 on either measure identified two unusually low-performing participants, both of whom were excluded. Because the average scores in the web sample were lower and more variable ($M = 841.5$, $SD = 214.4$) than those in the supervised desktop sample ($M = 911.3$, $SD = 166.6$), we used the desktop sample to set a cutoff for the web sample. We set the cutoff at 400, below the lowest participant-level average score in the desktop sample (443.3), and excluded the nine web participants whose average scores fell below it (Fig. S1).

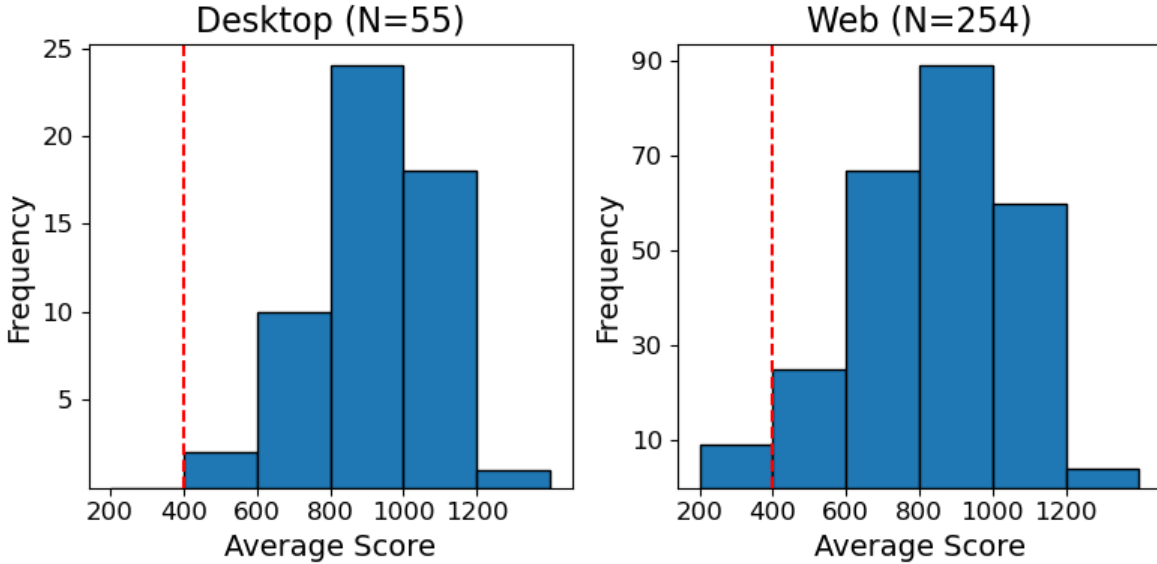


Figure S1. Distributions of participant-level average Crosswalk Task scores in the desktop ($N = 55$) and web ($N = 254$) samples before exclusion. The red dashed line indicates the average-score cutoff of 400. No desktop participants scored below this cutoff, whereas nine web participants did.

Second, web participants were screened for straightlining, which may indicate insufficient variation in questionnaire responses. Straightlining on the IGDS alone was not

treated as grounds for exclusion because selecting “never” for every item may constitute a valid response for participants without gaming experience or symptoms of gaming disorder. Accordingly, 11 web participants who straightlined on at least one non-IGDS questionnaire were excluded.

Finally, we screened both samples for unsuccessful fits of participant-specific AIRL models (Fu et al., 2018). Data from participants for whom a model could not be fitted reliably were excluded from analyses involving model-derived measures. Six participants in the web sample were excluded because discriminator accuracy remained above 90% despite repeated fine-tuning, indicating a failure to reach the prespecified fitting criterion. All participant-specific models in the desktop sample met the fitting criterion.

In total, two participants from the desktop sample ($N = 55$) and 25 from the web sample ($N = 254$) were excluded, with one participant in the web sample meeting multiple exclusion criteria. The final combined sample comprised 282 participants.

S4 Risk Preference Indicators

Based on data from the combined sample ($N = 282$), five behavioral measures and six IRL reward measures were selected as indicators of risk preference. The behavioral indicators comprised *Average Score*, *Episode Clear Ratio*, *Run Over Ratio*, *Road Crossing Success Ratio*, and *Crosswalk Searching*. The IRL reward indicators comprised *Safest Road Crossing*, *Waiting in Crosswalk for Car to Pass*, *High Risk but Empty Last Road*, *Approaching Crosswalk in Current Road*, *Approaching Next Crosswalks*, and *Moving Up Toward Car*.

All measures were calculated separately for each participant using data from the main sessions, with practice-session data excluded. Descriptive statistics for the 11 indicators are presented in Table S2.

Table S2. Descriptive statistics of risk preference indicators.

Indicator	<i>M</i>	<i>SD</i>	Min	Max
Average Score	877.360	187.374	444.444	1304.505
Episode Clear Ratio	0.535	0.121	0.079	0.778
Road Crossing Success Ratio	0.929	0.022	0.869	0.975
Crosswalk Searching	0.089	0.265	-1.000	1.000
Run Over Ratio	0.424	0.097	0.133	0.672
[IRL] Safest Road Crossing	0.460	0.287	-0.600	1.190
[IRL] Waiting in Crosswalk for Car to Pass	-0.311	0.484	-1.861	0.752
[IRL] High Risk but Empty Last Road	-0.206	0.256	-1.073	1.185
[IRL] Approaching Crosswalk in Current Road	0.476	0.230	-0.529	1.269
[IRL] Approaching Next Crosswalks	0.394	0.221	-0.725	0.981
[IRL] Moving Up Toward Car	0.382	0.233	-0.583	0.950

S4.1 Behavioral Indicators

The five behavioral indicators captured behavior at different levels of analysis. *Average Score*, *Episode Clear Ratio*, and *Run Over Ratio* were aggregate measures derived from episode-level outcomes, whereas *Road Crossing Success Ratio* and *Crosswalk Searching* were finer-grained measures derived from road-level outcomes and action-level events, respectively.

Average Score was calculated by dividing the total score accumulated during the experiment by the total number of episodes. Unlike *Total Score*, which increases with task duration, and *Best Score*, which reflects only the highest-scoring episode, *Average Score* summarizes performance across all episodes while reducing sensitivity to task duration. Although the theoretical lower and upper bounds of the measure were $-1,000$ and $2,700$, respectively, neither bound was attainable during ordinary task performance.

Three indicators were defined as proportions. *Episode Clear Ratio* was the proportion of episodes ending with the participant reaching the destination, *Run Over Ratio* was the proportion of episodes ending in a vehicle collision, and *Road Crossing Success Ratio* was the proportion of attempted road crossings in which the participant successfully reached the other side. All three had a theoretical range of 0 to 1.

Crosswalk Searching measured the extent to which each participant moved sideways

to search for a crosswalk on high-penalty roads relative to low-penalty roads. The measure was calculated as the normalized difference between the numbers of sideways movements made in the safe zone when no crosswalk was visible:

$$\frac{N_{\text{high}} - N_{\text{low}}}{N_{\text{high}} + N_{\text{low}} + 10^{-6}}.$$

For each participant, N_{low} and N_{high} represented the counts of these movements across all low- and high-penalty roads, respectively. The 10^{-6} term prevents division by zero, assigning a value of zero when neither count was positive. Values lie between -1 and 1 , with higher values indicating a greater tendency to search for a crosswalk on high-penalty roads than on low-penalty roads.

S4.2 IRL Reward Indicators

The six IRL reward indicators were computed from participant-specific reward functions inferred through AIRL. For each indicator, we pooled the main-session state-action trajectories of all participants in the combined sample ($N = 282$) and selected state-action pairs satisfying the indicator-specific conditions. We evaluated the same selected pairs with each participant’s inferred reward function and used the mean of the resulting reward values as that participant’s indicator value.

Safest Road Crossing captures *moving up* to cross a currently empty road using a crosswalk, thereby eliminating the possibility of a future collision penalty even though no vehicle poses an immediate threat (Fig. S2a). Its negative correlation with DOSPERT scores ($r = -.23$, $p < .001$) supports its interpretation as a risk-averse indicator. A higher indicator value indicates a greater inferred reward for this precautionary condition.

Waiting in Crosswalk for Car to Pass captures remaining stationary (*no-op*) inside a crosswalk when a vehicle is passing in front of the agent (Fig. S2b). It is also interpreted as a risk-averse indicator because of its negative correlation with DOSPERT scores ($r = -.23$,

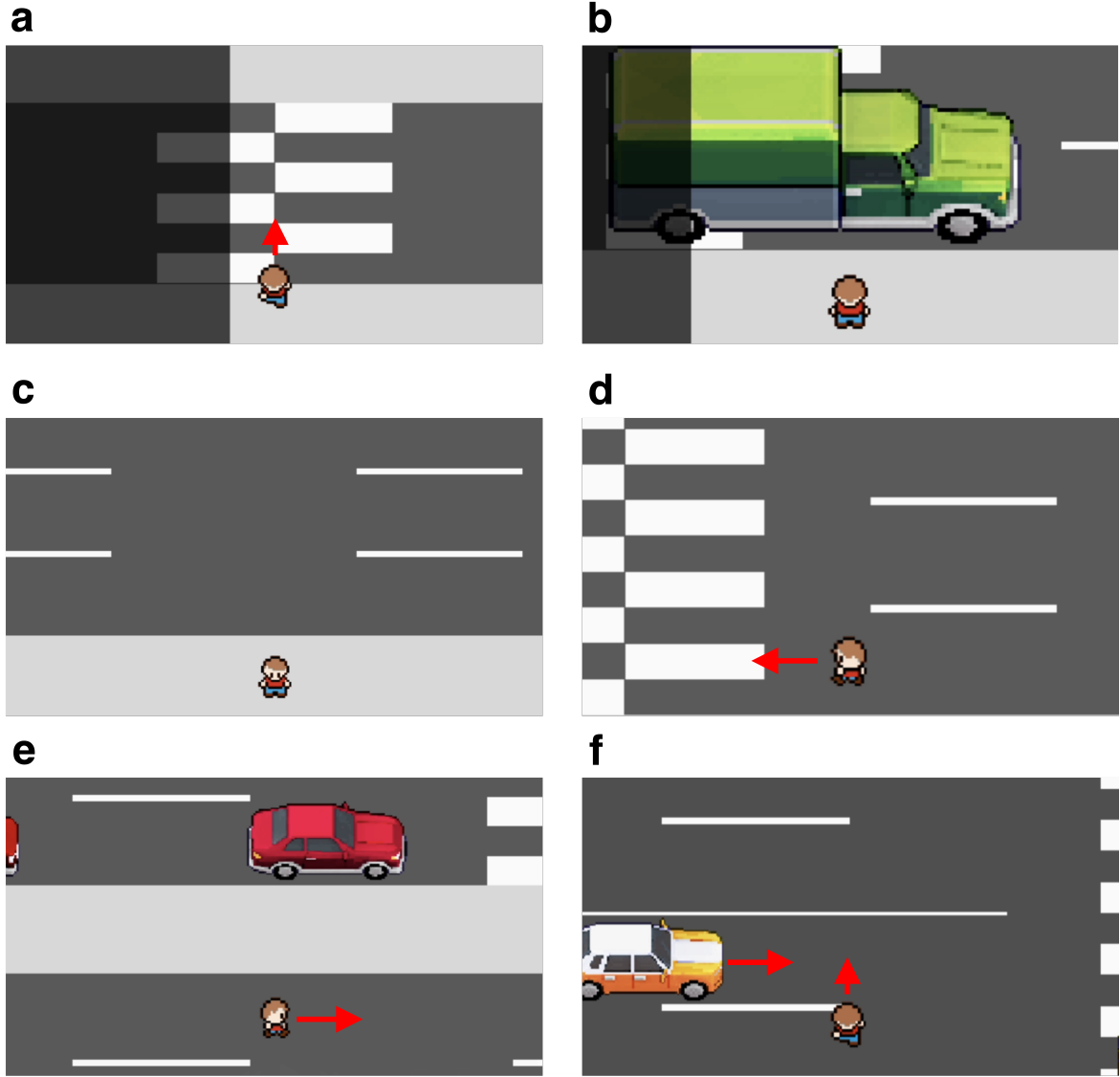


Figure S2. Visual examples of the conditions for each IRL reward indicator: (a) Safest Road Crossing; (b) Waiting in Crosswalk for Car to Pass; (c) High Risk but Empty Last Road; (d) Approaching Crosswalk in Current Road; (e) Approaching Next Crosswalks; and (f) Moving Up Toward Car.

$p < .001$). A higher indicator value indicates a greater preference for this precautionary condition.

High Risk but Empty Last Road captures a temporary opportunity to cross a high-risk, high-reward road. In this condition, the agent is in a safe zone facing the final road. No crosswalk or cars are visible, but red cars associated with a high collision penalty may appear at any moment (Fig. S2c). Successfully crossing the road would complete the episode

and earn a bonus, but doing so requires jaywalking and remains inherently risky: although the road appears clear at the outset, a car could appear while the agent is crossing. Its positive correlation with DOSPERT scores ($r = .19, p < .01$) supports its interpretation as a risk-taking indicator. A higher indicator value reflects a greater inferred reward for this condition, in which the agent has a temporary opportunity to pursue the episode-completion bonus despite the risk.

Approaching Crosswalk in Current Road captures *moving sideways* toward a visible crosswalk after entering a road where no cars are currently visible (Fig. S2d). A risk-averse strategy would be to return to the safe zone, approach and activate the crosswalk, and then re-enter the road. In contrast, moving directly toward the crosswalk while remaining in the road saves time but risks encountering a car that may suddenly appear. Its positive correlation with DOSPERT scores ($r = .18, p < .01$) supports its interpretation as a risk-taking indicator.

Approaching Next Crosswalks captures *moving sideways* toward visible crosswalks on subsequent roads while the agent is inside an empty road (Fig. S2e). Unlike those targeted in *Approaching Crosswalk in Current Road*, the target crosswalks are on roads the agent has not reached yet. A risk-averse strategy would be to focus on crossing the current road before considering the crosswalks on other roads. In contrast, moving toward the crosswalks while inside the current road risks encountering a car that may suddenly appear. Its positive correlation with DOSPERT scores ($r = .20, p < .001$) supports interpreting this event as risk taking.

Moving Up Toward Car captures *moving up* into the next lane when a vehicle is approaching in that lane (Fig. S2f). A risk-averse strategy would be to stay in the current lane until the car passes the agent. However, depending on the vehicle’s speed and distance, the agent may be able to finish crossing the lane before the vehicle arrives. Thus, moving up in such a condition would be considered a risk-taking indicator, with its positive correlation with DOSPERT scores ($r = .22, p < .001$) providing supporting evidence.

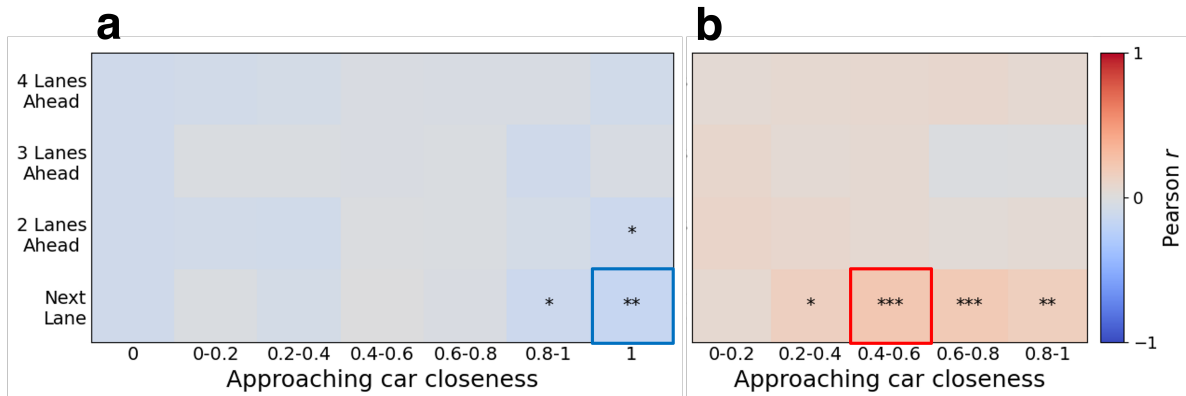


Figure S3. Selection of approaching-car proximity conditions for (a) *Waiting in Crosswalk for Car to Pass* and (b) *Moving Up Toward Car*. A value of 0 indicates that no vehicle is visible in the lane; a value of 1 indicates a vehicle immediately in front of the agent.

Various features from the 113-dimensional observation vector were considered when selecting and specifying the indicator conditions. *Waiting in Crosswalk for Car to Pass* requires a car immediately in front of the agent, a condition most strongly correlated with DOSPERT scores among similar candidates (Fig. S3a). *Moving Up Toward Car* uses a proximity range of 0.4–0.6, which showed the strongest correlation and places the approaching car roughly halfway between the edge of the map and the agent, making the safety of crossing more ambiguous than when the car is farther away or closer (Fig. S3b).

S5 Discriminant Validity

We evaluated the discriminant validity of the Crosswalk Task by examining the specificity of its associations with self-reported risk taking and comparing its correlations with DOSPERT scores to those of traditional behavioral tasks.

S5.1 Construct Specificity

Figure S4 shows the correlations of each risk preference indicator with BIS-11 total and subscale scores and STAI state and trait anxiety scores. Only two correlations were significant: scores on the BIS-11 Motor subscale were correlated with *Run Over Ratio* ($r = .12, p < .05$)

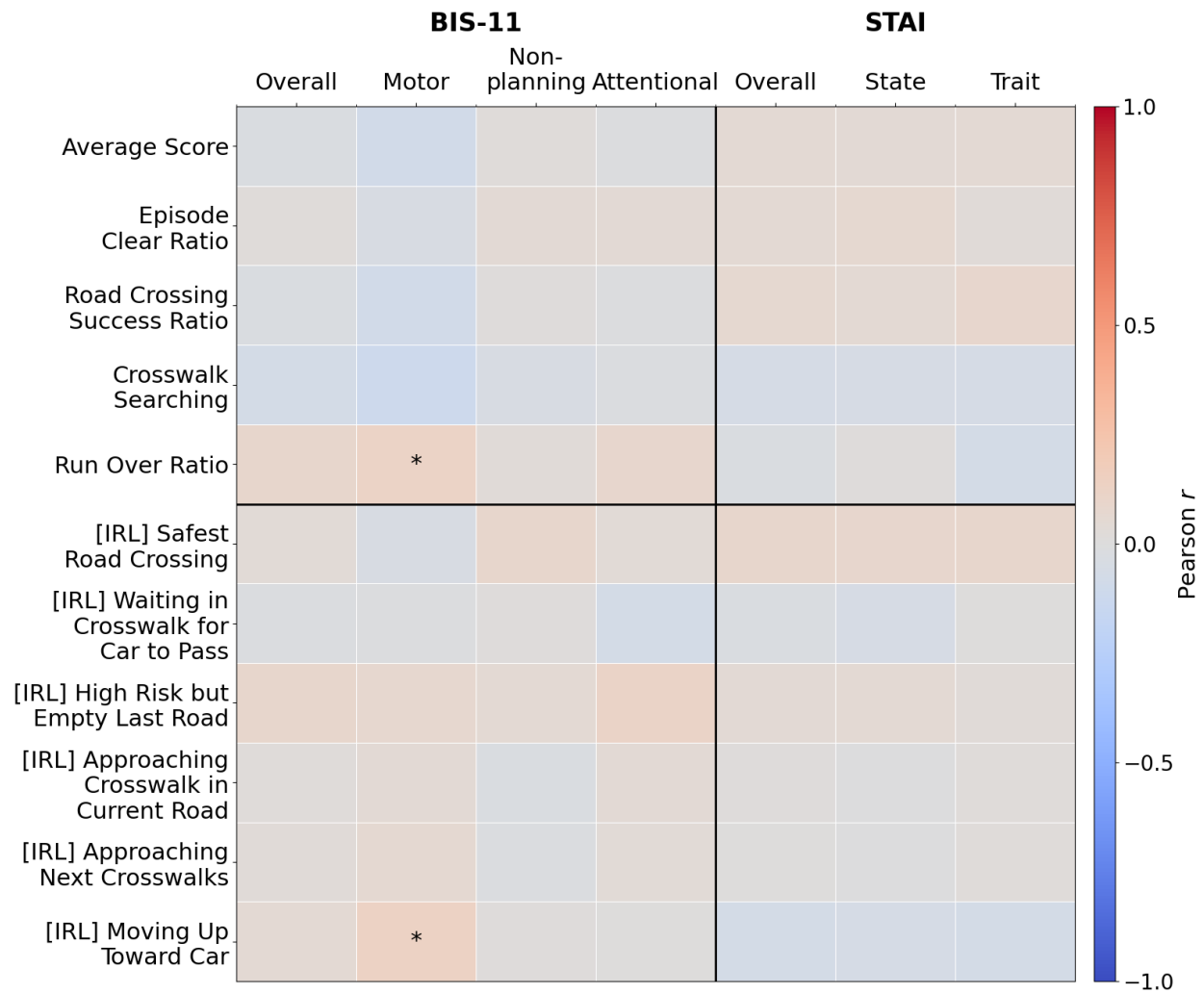


Figure S4. Pearson correlations of Crosswalk Task measures with BIS-11 total and subscale scores and STAI state and trait anxiety scores. * $p < .05$.

and *Moving Up Toward Car* ($r = .13, p < .05$). No other associations were significant.

S5.2 Comparison With Traditional Tasks

In addition to the Crosswalk Task, participants in the desktop sample completed the Delay Discounting Task (DDT; Green and Myerson, 2004; Mazur, 2013) and the Choice under Risk and Ambiguity Task (CRA; Levy et al., 2010). The DDT yielded a log-transformed discounting-rate parameter ($\log k$), with higher values indicating steeper discounting of delayed rewards. The CRA yielded a risk-attitude parameter (α), with higher values indicating greater risk preference. We included the CRA risk-attitude parameter as a direct measure of risk preference and the DDT discounting-rate parameter as a related decision-making measure, given its theoretical and empirical links to risk preference (Johnson et al., 2020).

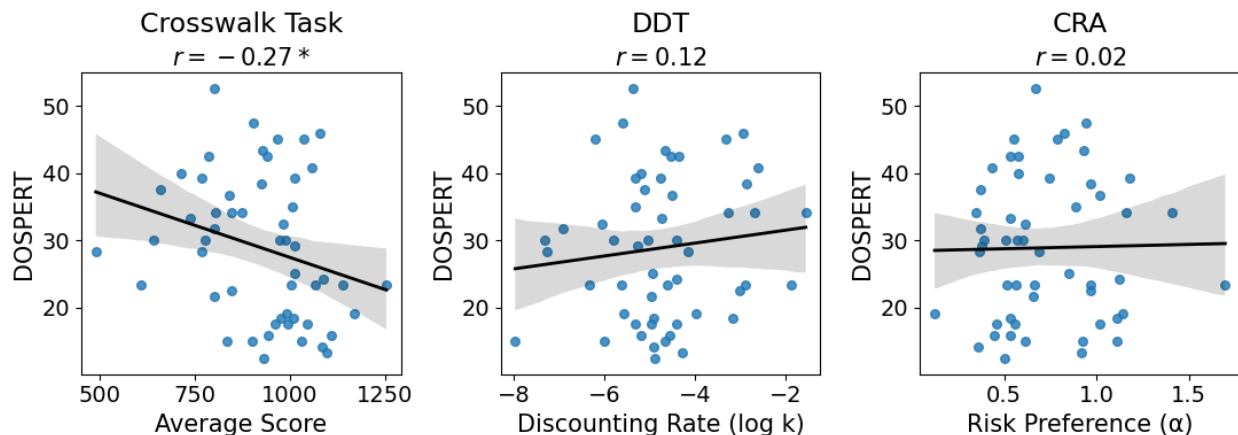


Figure S5. Pearson correlations of the DOSPERT total score with the Crosswalk Task *Average Score*, the Delay Discounting Task (DDT) discounting rate ($\log k$), and the Choice under Risk and Ambiguity Task (CRA) risk-attitude parameter (α) in the desktop sample. The shaded areas indicate 95% confidence intervals. $^*p < .05$.

Figure S5 shows that neither the DDT discounting-rate parameter ($r = .12, p > .05$) nor the CRA risk-attitude parameter ($r = .02, p > .05$) was significantly correlated with the DOSPERT total score. In contrast, several Crosswalk Task measures showed significant correlations with the DOSPERT total score, including *Average Score*, which is shown in the figure.

S6 Regression Analysis

S6.1 Elastic Net

Elastic net regression (Zou and Hastie, 2005) was performed using the `glmnet` package (Friedman et al., 2010) and the `easym1` package (Ahn et al., 2017). The mixing parameter was set to $\alpha = 0.5$, assigning equal weight to L1 (lasso) and L2 (ridge) regularization. The regularization path was configured with 100 candidate λ values and a minimum-to-maximum λ ratio of 10^{-4} . Five-fold cross-validation was used to tune the regularization parameter based on mean squared error (MSE).

The dataset was divided into training and held-out test sets using an 80:20 split, with 1,000 random divisions and 100 prediction iterations per division. The random seed was set to 42. Predictive performance was evaluated using Pearson’s correlation coefficient (r) between predicted and observed DOSPERT scores. Table S3 summarizes feature selection and predictive performance for the three models.

Table S3. Elastic net performance.

	Model 1	Model 2	Model 3
Feature Selection	Behavioral: 3/5	IRL: 6/6	IRL: 6/6 Behavioral: 2/5
Pearson r (Train)	0.2777	0.3641	0.4238
Pearson r (Test)	0.2313	0.3058	0.3553

S6.2 Multilayer Perceptron

A multilayer perceptron (MLP) was trained to predict DOSPERT total scores using 11 risk preference indicators as input features. The network consisted of two hidden layers with two hidden units and a rectified linear unit (ReLU) activation function. The model was optimized using Adam with a learning rate of 1×10^{-4} , a batch size of 16, and a weight

decay of 1×10^{-5} . MSE was used as the loss function.

The dataset was randomly divided into training and held-out test sets using an 80:20 split with five different random seeds (42–46). For each split, five-fold cross-validation was performed within the training set, with shuffling enabled and the same random seed used to generate the folds, resulting in 25 model fits in total. Model performance was evaluated using Pearson’s correlation coefficient (r), the coefficient of determination (R^2), and root mean squared error (RMSE). Performance metrics were summarized across the 25 model fits as the mean $\pm$ standard deviation.

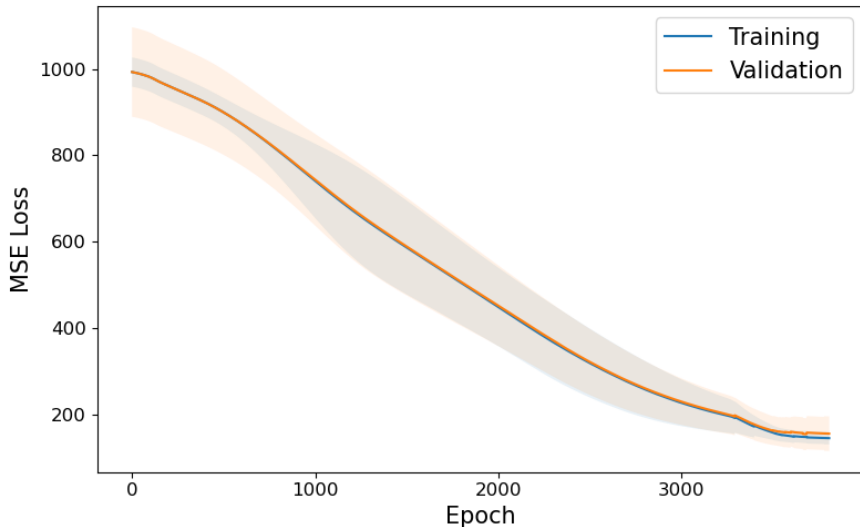


Figure S6. Training and validation loss curves for the MLP across 25 model fits. Solid lines represent the mean MSE loss at each epoch, and shaded areas indicate one standard deviation above and below the mean.

Training was performed for a maximum of 50,000 epochs with early stopping based on the validation loss. Early stopping was implemented with a patience of 50 epochs and a relative improvement threshold of 10^{-4} , and the model parameters corresponding to the best validation loss were restored after training. Figure S6 shows the mean training and validation loss curves across the 25 model fits.

Table S4 summarizes predictive performance across the 25 model fits. Mean Pearson correlations and RMSE values were similar between the validation and held-out test sets. However, the mean test-set R^2 was low (0.0375), indicating limited predictive accuracy

despite the positive correlation between predicted and observed DOSPERT scores.

Table S4. MLP performance.

	Significant runs	Pearson r	R^2	RMSE
Train	100%	$.4067 \pm 0.0483$	0.1487 ± 0.0534	12.0429 ± 0.6271
Val	68%	$.3500 \pm 0.1461$	0.1063 ± 0.0913	12.1663 ± 1.2979
Test	68%	$.3392 \pm 0.1407$	0.0375 ± 0.1498	12.2229 ± 1.1333

Note. Values are reported as the mean $\pm$ SD across 25 model fits. The *Significant runs* column reports the percentage of model fits in which the Pearson correlation between predicted and observed scores was statistically significant ($p < .05$).

S7 Clustering Analysis

We selected the number of clusters for k -means clustering (McQueen, 1967) using the mean silhouette score and the inertia curve. The silhouette score measures within-cluster cohesion relative to separation from the nearest other cluster, with higher values indicating better-defined clusters (Rousseeuw, 1987). Inertia is the sum of squared Euclidean distances between observations and their assigned centroids. Because inertia generally decreases as k increases, the elbow criterion identifies where adding more clusters yields diminishing reductions in inertia (Thorndike, 1953). Among the evaluated solutions ($k = 2-7$), $k = 2$ yielded the highest mean silhouette score (0.209; Fig. S7a), while the inertia curve suggested an elbow near $k = 2$ (Fig. S7b), supporting this choice.

The final cluster assignments were obtained by fitting k -means to the 11 risk preference indicators derived from the Crosswalk Task, with $k = 2$, 10 initializations, and a fixed random seed of 42. DOSPERT scores were excluded from both cluster-number selection and cluster assignment and were subsequently used for external validation and labeling of the resulting groups as *Risk Taking* and *Risk Averse*. Figure S7c visualizes the behaviorally derived groups along three of the 11 input indicators. This visualization illustrates how the groups differ in their task behavior, providing an interpretable view of the clustering

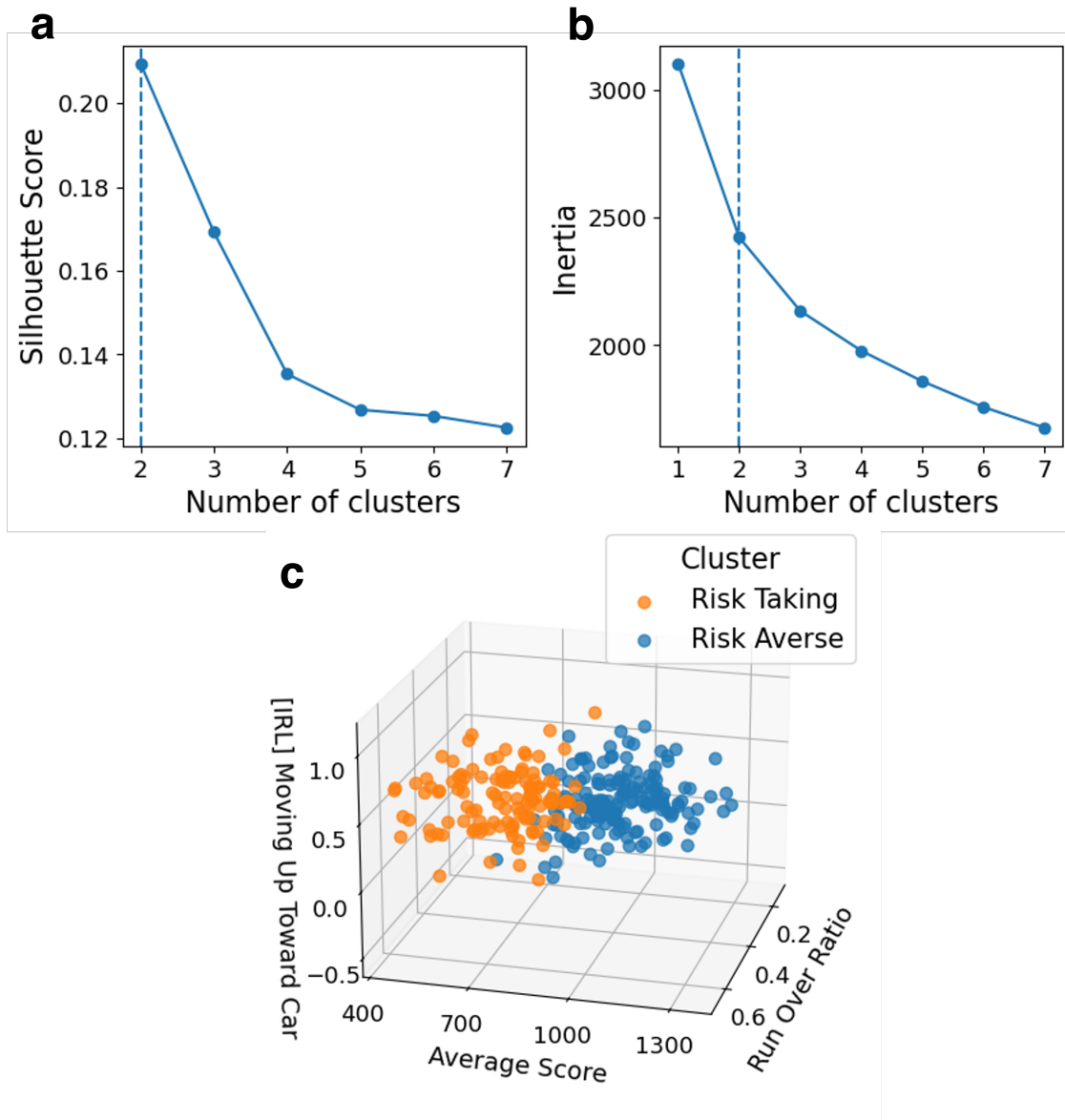


Figure S7. Selection of the number of clusters and results of k -means clustering. (a) Mean silhouette scores for $k = 2-7$. (b) Within-cluster sum of squared distances (inertia) for $k = 1-7$. (c) Three-dimensional visualization of the clusters derived from all 11 Crosswalk Task indicators, plotted using three of the input indicators: *Average Score*, *Run Over Ratio*, and *Moving Up Toward Car*. Cluster labels were assigned based on DOSPRT scores after clustering.

obtained using all 11 indicators.

References

- Ahn, W.-Y., Hendricks, P., & Haines, N. (2017). Easymml: Easily build and evaluate machine learning models. *BioRxiv*, 137240.
- 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.
- Cho, S. H., & Kwon, J.-H. (2017). A validation of the korean version of the internet gaming disorder scale (k-igds): Findings from a community sample of adults. *Korean Journal of Clinical Psychology*, 36(1), 104–117.
- Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33, 1–22.
- Fu, J., Luo, K., & Levine, S. (2018). Learning robust rewards with adversarial inverse reinforcement learning. *International Conference on Learning Representations (ICLR)*.
- Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological bulletin*, 130(5), 769.
- Gyeong, H., Lee, H.-K., & Lee, K. (2012). Factor analysis of the young’s internet addiction test: In korean college students group. *Journal of the Korean Neuropsychiatric Association*, 51(1), 45–51.
- Johnson, K. L., Bixter, M. T., & Luhmann, C. C. (2020). Delay discounting and risky choice: Meta-analytic evidence regarding single-process theories. *Judgment and Decision making*, 15(3), 381–400.
- Lemmens, J. S., Valkenburg, P. M., & Gentile, D. A. (2015). The internet gaming disorder scale. *Psychological assessment*, 27(2), 567.
- Levy, I., Snell, J., Nelson, A. J., Rustichini, A., & Glimcher, P. W. (2010). Neural representation of subjective value under risk and ambiguity. *Journal of neurophysiology*, 103(2), 1036–1047.

- Mazur, J. E. (2013). An adjusting procedure for studying delayed. *The Effect of Delay and of Intervening Events on Reinforcement Value: Quantitative Analyses of Behavior, Volume V*, 55.
- 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.
- Ng, A. Y., Russell, S., et al. (2000). Algorithms for inverse reinforcement learning. *Icml*, 1(2), 2.
- 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.
- Rodd, J. M. (2024). Moving experimental psychology online: How to obtain high quality data when we can't see our participants. *Journal of Memory and Language*, 134, 104472.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, 53–65.
- 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.
- Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18(4), 267–276.
- Young, K. (2016). *Internet addiction test (iat)*. Stoelting.
- 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.