

# What makes an action sequence enjoyable to watch?

Jean-Peïc Chou<sup>\*1</sup>, Kristine Zheng<sup>2</sup>, Junyi Chu<sup>2</sup>, Maneesh Agrawala<sup>1</sup> & Judith E. Fan<sup>1,2</sup>

<sup>1</sup>Department of Computer Science, Stanford University, United States

<sup>2</sup>Department of Psychology, Stanford University, United States

\*jeanpeic@stanford.edu

## Abstract

People often seek out ways to watch others perform complex action sequences (e.g., sports). What makes some sequences more enjoyable to watch than others? We generated 24 video clips of gameplay from a Flappy Bird-style video game. Clips varied in difficulty (how often players succeeded on average) and in moment-to-moment uncertainty (how likely the player was to crash at any given step). Participants ( $N = 864$ ) rated each video on one of three dimensions: how much they enjoyed it, how difficult the level appeared, or how dangerous the player's trajectory appeared. We found that participants preferred videos where the player seemed to be completing more difficult obstacle courses, but dangerousness did not predict enjoyment ratings. These findings show how procedurally generated stimuli can isolate the factors that affect how enjoyable an action sequence is to watch.

**Keywords:** affect; motivation; emotion; uncertainty; dynamics

## Introduction

On January 25, 2026, over six million people tuned in on Netflix to watch Alex Honnold ascend the 1,667-foot Taipei 101 tower—bare-handed, without ropes, and in real time. Spectators weren't learning to climb themselves, nor did the outcome affect their lives in any practical way. Rather, they watched for enjoyment. What makes some performances more enjoyable to watch than others?

Survey studies have found that spectators enjoy watching "close contests," "amazing feats," and "skillful plays" (Pirouz et al., 2015; Pizzo et al., 2018; Zillmann et al., 1989). These observations suggest a variety of features that may explain enjoyment, including how difficult a task is or how close to failure the performer seems moment to moment. However, it remains unclear what information observers use to form these enjoyment judgments, whether such judgments are shared across viewers, and whether they track measurable properties of the events themselves.

One framework linking affective responses to stimulus features comes from appraisal theories of emotion, which propose that feelings arise from how people interpret events (Ellsworth and Scherer, 2003). Such appraisals can range from simple evaluations, such as stimulus novelty or uncertainty, to more complex judgments about personal relevance or threat (Moors et al., 2013; Smith and Kirby, 2012). This framework predicts that how observers interpret an action should affect what they feel while watching it.

A long history of research in cognitive science has found

that action understanding involves attending to both the static features of an environment and the moment-to-moment dynamics of actions (Dennett, 1987; Heider and Simmel, 1944; Wellman, 2002). Adults, children, and infants are sensitive to the physical constraints that increase action costs (e.g., climbing steeper hills; Gergely and Csibra, 2003; Liu et al., 2017) and danger (e.g., jumping over deeper trenches; Gjata et al., 2022; Liu et al., 2022), and they expect agents to minimize such costs. People also use differences in action trajectories to infer others' goals, beliefs, and competence (Jara-Ettinger et al., 2016, 2020; Liu et al., 2026). For example, inefficient behaviors may signal ignorance (Baker et al., 2009) or incompetence (Leonard et al., 2019). These evaluations extend beyond social inferences. Reasoning about external constraints and action dynamics also informs judgments about task difficulty (Gweon et al., 2017; Yildirim et al., 2019) and interestingness (Holdaway et al., 2021; Martinez et al., 2023).

Here, we investigate whether the same dimensions that observers use to understand actions also predict their subjective feelings of enjoyment. As a first step, we consider two salient features: how challenging the environment is, and the moment-to-moment dangerousness of the agent's trajectory.

We developed and tested several competing hypotheses for how difficulty and dangerousness might affect how much viewers enjoy a performance. First, viewers might enjoy performances situated in more difficult environments. When success seems improbable, observers may derive pleasure from witnessing novel, unlikely outcomes (Pirouz et al., 2015) or from appreciating the skill required to overcome challenges (Barney and Pennington, 2023; Pizzo et al., 2018). Second, viewers may prefer performances with greater moment-to-moment uncertainty, as when agents execute risky maneuvers or navigate precarious situations. While many theoretical accounts treat the experience of uncertainty as inherently aversive (Madrigan et al., 2011), observing the resolution of such momentary dangers may be rewarding (Kaspar et al., 2016; Moulard et al., 2019). Third, the two factors may interact: dangerous maneuvers may be especially enjoyable to watch in high-difficulty environments, where such danger is necessary for success, but reduce enjoyment in low-difficulty settings. When safer paths are available, unnecessarily dangerous movements may suggest avoidable errors and imply lower skill, thus making the performance less enjoyable to watch. Finally, uncertainty about agent trajectories may not specifically increase enjoyment but instead amplify emotional

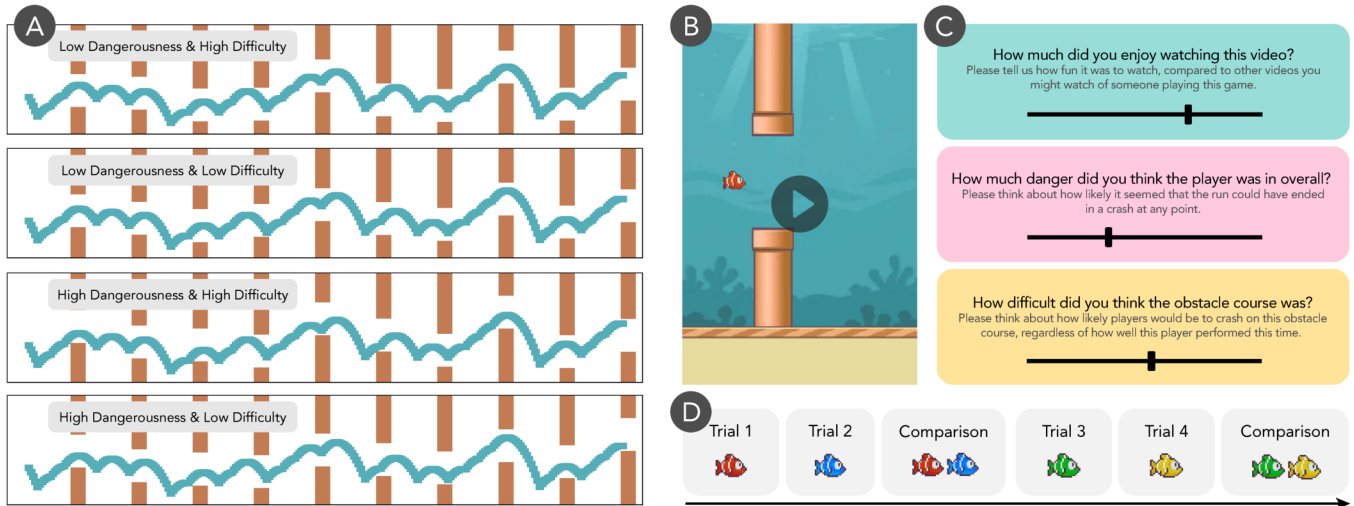


Figure 1: (A) Schematic visualizations of four stimuli showing the key experimental manipulation (Map difficulty  $\times$  Dangerousness). Each example shows the same base trajectory (in blue, moving from left to right) in different obstacle layouts (in brown). Difficulty and dangerousness scores were computed using agent-based simulations (see Methods). Obstacle courses with higher difficulty scores have a higher rate of failure across simulated trajectories. Higher dangerousness scores reflect trajectories with smaller margins for error, meaning they would be more likely to result in collision under noisy perception or imprecise motor control. (B) A single video frame. (C) Participants rated each video on one of three dimensions using a 0–100 slider. The type of rating task was manipulated between groups of participants. (D) Study procedure. Participants viewed four gameplay videos (one per Dangerousness  $\times$  Map difficulty condition) and rated each for either enjoyment, dangerousness, or difficulty.

intensity, regardless of valence (Bar-Anan et al., 2009). On this account, more dangerous trajectories should elicit more intense emotional responses — positive or negative — and produce greater spread or bimodality in self-reported enjoyment. While these four hypotheses are not mutually exclusive, they make different predictions about the relative contributions of difficulty and dangerousness to enjoyment.

We test these hypotheses in a video game environment (Figure 1) modeled after Flappy Bird, a simple single-player game that has attracted millions of online spectators (Dredge, 2015; Google, 2014). We generated gameplay videos by simulating player agents, allowing us to independently manipulate environment difficulty and trajectory dangerousness while holding visual features and outcomes constant (Figure 1A). Participants viewed these videos and rated each on apparent difficulty, apparent dangerousness, or their own enjoyment.

This design lets us directly test how difficulty and dangerousness influence subjective enjoyment. Our analysis focuses on three questions: (1) the extent to which observers agree in their judgments of enjoyment, difficulty, and dangerousness; (2) the degree to which subjective judgments of dangerousness and difficulty align with model-based estimates derived from simulated agent behavior; and most importantly, (3) the extent to which apparent dangerousness and apparent difficulty predict enjoyment, and whether these features interact.

## Methods

### Participants

We recruited 905 US adults from Prolific (*age*: median = 40; *gender*: 44.2% female, 53.7% male). We excluded 41 participants for incomplete test trials, leaving a final analyzed sample of 864 participants, evenly assigned across the dangerousness, map difficulty, and enjoyment rating conditions ( $N = 288$  per condition). The study took approximately five minutes to complete, and participants received \$1.50 USD in compensation.

### Stimuli

We created 13-second videos from a custom variant of the game Flappy Bird, starting with a 3-second countdown followed by 10 seconds of gameplay footage. In this game, players navigate a fish through vertical obstacles ("pipes") by executing discrete upward "jumps", while facing constant downwards gravity and constant horizontal velocity. A run ends if the fish collides with an obstacle pipe, the ground, or the top of the screen.

To control for outcome, all videos show the fish successfully navigating exactly 10 consecutive obstacles. We generated six unique trajectories and created four videos with each trajectory, following a  $2 \times 2$  factorial design (low/high dangerousness  $\times$  low/high difficulty). This design lets us decouple dangerousness from difficulty while controlling for trajectory shape (see Figure 1A).

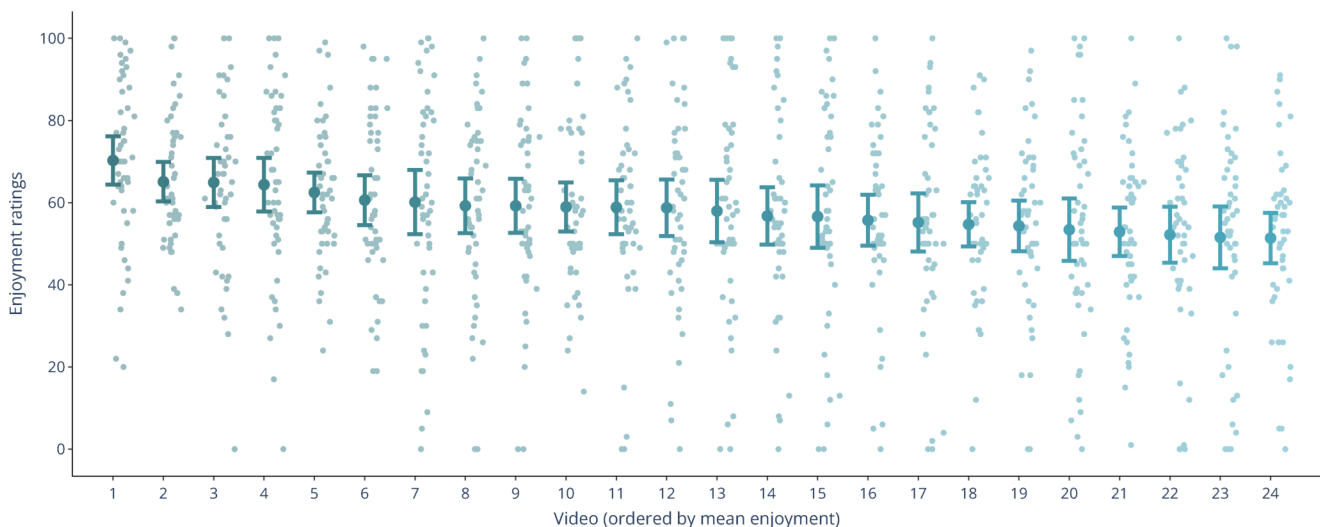


Figure 2: Distribution of enjoyment ratings for each of the 24 videos from highest to lowest mean enjoyment rating. Larger dots and error bars show the mean and 95% CI; smaller dots show individual ratings (48 per video).

**Operationalizing difficulty and dangerousness.** Our goal was to independently manipulate two stimulus features: how difficult the environment is overall, and the moment-to-moment dangerousness of an agent’s trajectory. Both features are defined relative to an agent’s abilities. For example, a player with slower reaction time may be more likely to collide with nearby obstacles. Because both features depend on agent ability, we quantified them by aggregating performance statistics across simulated agents with varying sensorimotor abilities.

We trained three reinforcement-learning agents to play the game under different levels of sensorimotor impairment (Low, Medium, High). Impairment combined perceptual noise (i.e., Gaussian noise on the perceived position of upcoming pipes) and motor noise (i.e., variable jump execution delay, sampled from a normal distribution), with both increasing across impairment levels. Agents were trained using Proximal Policy Optimization (PPO; Schulman et al., 2017) to maximize survival time, yielding control policies adapted to their sensorimotor limitations.

We operationalized **dangerousness** as susceptibility to imminent failure. For each state in a trajectory, we queried the agent’s learned value function, which estimates expected remaining survival time. Because our reward structure equates survival with cumulative reward, low values indicate states from which the agent does not expect to survive long. We averaged the negated value across the full trajectory so that higher scores indicate greater dangerousness.

We operationalized **difficulty** as the probability of failing to complete an obstacle course (Gudmundsson et al., 2018; Isaksen et al., 2017). For each obstacle layout, we estimated failure probability from 1,000 simulated runs per agent, then averaged across all three agents.

**Generating videos.** Using these dangerousness and difficulty scores, we created 24 videos following a factorial design, systematically varying difficulty and dangerousness while controlling for trajectory shape and outcome (Figure 1A). All videos showed successful completion of exactly 10 obstacles.

We first identified six base trajectories with distinct movement patterns. From 1,000 simulated runs per agent, we selected the trajectories with lowest and highest mean vertical movement amplitude within each agent (3 agents  $\times$  2 amplitudes = 6 trajectories). This provided diversity in movement dynamics.

For each base trajectory, we generated potential obstacle layouts that varied in difficulty and dangerousness. We sampled 1,000 collision-free layouts by randomly varying pipe heights around the trajectory. This established the feasible range of dangerousness and difficulty scores for each trajectory. We then selected four layouts per trajectory using extreme percentile thresholds ( $\leq$  5th,  $\geq$  95th), creating a 2 $\times$ 2 factorial design (low/high dangerousness  $\times$  low/high difficulty).

## Procedure

Participants first read instructions explaining the game rules and watched a single example video — gameplay in an extremely easy environment (i.e., large obstacle spacing) — to familiarize themselves with the dynamics. They were then randomly assigned to one of three rating conditions (dangerousness, difficulty, or enjoyment), each with condition-specific instructions (Figure 1C).

Each participant viewed four videos in sequence, one from each stimulus condition (high/low dangerousness  $\times$  high/low difficulty). Videos were presented in two blocks of two. The two videos within a block shared a base trajectory but were

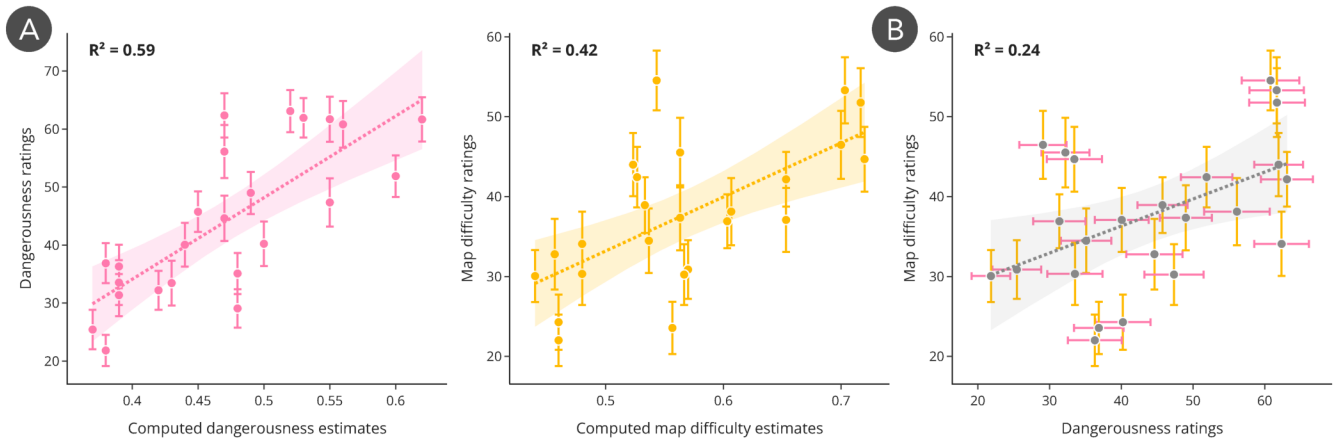


Figure 3: (A) Correlations between ratings and model-based estimates for dangerousness and map difficulty. Points show per-video mean ratings. Error bars show 95% bootstrapped CIs. Best fit lines show linear estimate with 95% CI. (B) Correlation between map difficulty and dangerousness ratings.

rendered on different obstacle courses, and together the four videos spanned all four (dangerousness  $\times$  difficulty) conditions. Each video featured a different colored fish (red, blue, green, and yellow) to help participants distinguish between them. The assignment of stimulus condition to video order and the mapping of fish colors to video order were both randomized across participants.

After watching each video, participants rated it on their assigned dimension using a continuous slider (0–100). Following each pair of videos, participants completed a forced-choice comparison, indicating which of the two videos scored higher on their assigned dimension. After completing all ratings, participants answered a brief post-study survey in which they explained their judgments.

## Results

### Viewers reliably judge enjoyment, difficulty, and dangerousness

We first assessed whether videos elicited consistent judgments across participants. We assessed stimulus-level reliability using a permutation-based split-half procedure with Spearman–Brown correction over 10,000 random splits (Parsons, 2021).

Participants showed strong agreement when judging apparent dangerousness ( $r_{SB} = 0.93$ , 95% CI [0.87, 0.96]) and map difficulty ( $r_{SB} = 0.83$ , 95% CI [0.71, 0.91]). Participants showed moderate agreement in enjoyment ratings ( $r_{SB} = 0.57$ , 95% CI [0.25, 0.78]). Per-video enjoyment ratings appeared broadly unimodal, with disagreement expressed as spread rather than clear separation into groups (mean SD across videos = 22.3). Nevertheless, enjoyment ratings varied across stimuli with mean ratings ranging from 51.4 to 70.3 (see Figure 2).

### Difficulty and dangerousness judgments track model-based estimates, but difficulty judgments are also influenced by dangerousness

Next, we assessed whether self-reported judgments aligned with computed estimates derived from agent-based models.

Dangerousness ratings closely tracked dangerousness estimates ( $r(22) = 0.79$ ,  $p < 0.001$ ), indicating that on average, the judgments aligned with estimated susceptibility to failure (see Figure 3A). Forced-choice comparisons within participant provided converging evidence. Forced-choice judgments of which player was in more danger were near chance when both clips shared the same dangerousness condition (high-dangerousness & high-difficulty vs. high-dangerousness & low-difficulty: 53% vs 47%; low-dangerousness & high-difficulty vs. low-dangerousness & low-difficulty: 51% vs 49%;  $N = 96$  each). When a high-dangerousness clip was paired with a low-dangerousness clip ( $N = 384$ ), participants chose the high-dangerousness clip on 80% of trials.

Mean map difficulty ratings also correlated strongly with computed difficulty estimates ( $r(22) = 0.64$ ,  $p < 0.001$ ; see Figure 3A). However, difficulty judgments were also significantly associated with dangerousness, correlating with both mean dangerousness ratings ( $r(22) = 0.46$ ,  $p < 0.05$ ; see Figure 3B) and computed dangerousness estimates ( $r(22) = 0.61$ ,  $p < 0.01$ ). Pairwise choices within participants corroborated these results: when high-dangerousness & low-difficulty videos were paired with low-dangerousness & high-difficulty videos ( $N = 96$ ), participants selected each clip as “more difficult” about equally often (51% vs 49%). In comparison, aggregating across all pairwise matchups, high-dangerousness & high-difficulty and low-dangerousness & low-difficulty videos were consistently judged most and least difficult, respectively. This pattern suggests that both obstacle layout and trajectory-level dangerousness shaped difficulty judgments.

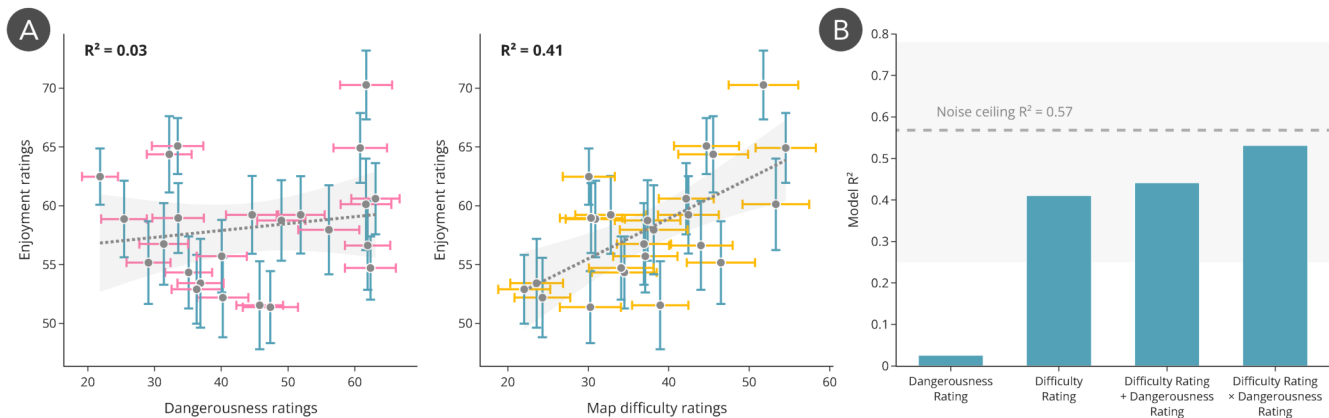


Figure 4: (A) Correlations between enjoyment and dangerousness and map difficulty ratings. Each point represents the mean rating across participants for one video. Error bars show 95% bootstrapped CIs. Best fit lines show linear estimate with 95% CI. (B) Amount of variation in enjoyment that is attributable to ratings and trajectory. Y-axis show adjusted  $R^2$  values from linear mixed effects models with different predictor terms, indicated on the X-axis. Noise ceiling shows maximal achievable  $R^2$ , estimated from the observed between-participant consistency in enjoyment ratings. Grey area indicates the noise ceiling 95% CI.

### Viewer enjoyment ratings track judgments of map difficulty but not dangerousness

Having established that judgments of dangerousness and map difficulty are reliable across viewers and track their computed estimates, we next tested whether these judgments predict enjoyment.

Videos reported as more difficult were also rated as more enjoyable ( $r(22) = 0.68$ ,  $p < 0.001$ , 95% CI = [0.32, 0.83]). By contrast, dangerousness ratings showed no relationship with enjoyment ( $r(22) = 0.11$ ,  $p = 0.62$ ; see Figure 4A). A multiple regression using both factors to predict average enjoyment across the 24 videos<sup>1</sup> explained almost half the observed variance ( $R^2 = 0.44$ ). Difficulty rating was a strong positive predictor of enjoyment, whereas adding dangerousness rating as a main effect did not significantly improve fit over difficulty alone (nested model comparison:  $F(1, 21) = 1.18$ ,  $p = 0.29$ ; see Figure 4B). Adding a difficulty  $\times$  dangerousness interaction term increased model fit ( $R^2 = 0.53$ ) without reaching statistical significance (nested ANOVA:  $F(1, 20) = 3.84$ ,  $p = 0.06$ ). We observed a positive interaction ( $\beta = 0.34$ ,  $p = 0.06$ ), suggesting that the enjoyment boost from higher difficulty was larger when dangerousness was also higher. We treat this interaction as exploratory given the marginal  $p$ -value and the small stimulus set ( $N = 24$ ).

Next, we examined whether apparent dangerousness influenced the spread of enjoyment responses. We found no evidence for such an effect: dangerousness ratings were only weakly correlated with the per-video standard deviation of enjoyment ratings, and this correlation was not statistically

<sup>1</sup>We also tested mixed-effects models, but with only four ratings per participant, the data structure did not support stable estimation of participant-level random effects.

significant ( $r(22) = 0.23$ ,  $p = 0.29$ ).

On average, viewers derived more enjoyment from performances on tasks they judged to be difficult, while apparent dangerousness had no consistent effect on enjoyment.

### Discussion

People often enjoy watching others in action, but the event features that drive this enjoyment are not well understood. Appraisal theories of emotion suggest that subjective feelings depend on how individuals evaluate the events they experience, while a long history of work on action understanding has shown that people readily extract information about both static environmental constraints and the moment-to-moment dynamics of agents' actions. Integrating these perspectives, we identified two candidate event features: the overall likelihood that an agent will succeed in a given environment, and the agent's susceptibility to failure at any given moment along a trajectory. We tested how people respond to these features using simulated gameplay videos that independently manipulated environment difficulty and trajectory-level dangerousness.

We found that video enjoyment ratings were predicted by environment difficulty, with no consistent effect of trajectory-level dangerousness. This pattern suggests that viewers primarily derive enjoyment from performances that overcome difficult environmental constraints, rather than from the moment-to-moment dangers agents encounter. Notably, this pattern held even though dangerousness influenced difficulty judgments themselves, suggesting that viewers extract and weight environmental difficulty independently of dangerousness when judging enjoyment.

The null effect of dangerousness is striking given that nearly

half of participants spontaneously mentioned dangerousness-related factors in a post-study survey explaining their responses. One possibility is that dangerousness mattered to some degree but was masked by measurement noise due to individual differences. For example, people differ in their general propensity to feel positive emotions (Hamilton et al., 1984) and in their tolerance for risk and uncertainty (Weber et al., 2002). Participants may also interpret “enjoyment” in diverse ways (e.g., as excitement, interest, or suspense), producing heterogeneous responses to the same stimuli. Consistent with this account, inter-rater reliability was lower for enjoyment than for difficulty or dangerousness judgments. Future work should collect more ratings per participant across a broader set of evaluative dimensions, which would help separate individual differences in trait-level enjoyment from inconsistency in how “enjoyment” is interpreted.

However, the contrast between participants’ explicit explanations and our experimental results may reflect more than just measurement noise. Instead, the contrast may reflect a substantive dissociation: people’s intuitions about what drives their enjoyment may diverge from the features that actually do. This dissociation points to the limits of introspection as a guide to aesthetic preference, and to the need for measures of enjoyment beyond self-report. In particular, future research should complement self-reports with more implicit measures of enjoyment, including physiological and behavioral measures of emotion and attention (e.g., pupillometry), and consider continuous measures to distinguish retrospective evaluations from enjoyment as it unfolds.

In this work, we chose the Flappy Bird environment because the constrained action space provided a controlled setting for dissociating environment-based difficulty and moment-to-moment trajectory dangerousness. However, the correlations between perceived difficulty and dangerousness suggest that these factors are not independent appraisals. Viewers may use inferences about an agent’s competence and intentionality when judging both the difficulty of the obstacle course and the dangerousness of the agent’s actions. For example, a dangerous-looking trajectory may make an otherwise easy layout appear difficult if a competent player genuinely seems to struggle. Distinguishing these interpretations requires environments in which players can pursue different trajectories through the same situation, which would allow observers to infer whether dangerousness reflects task demands, intentional risk-taking, or lack of competence. Our Flappy Bird environment is limited in this respect because the obstacle layouts strongly constrain the range of viable trajectories. Increasing the range of possible trajectories and playing styles, for example by widening obstacle gaps or leveraging games with richer action spaces, would make it easier to distinguish these interpretations.

Another limitation of the environment concerns potential confounds between task difficulty and trajectory shape. Because difficulty is defined with respect to obstacle positions, larger height changes between successive pipes both increased

difficulty and required agents to make larger vertical movements. Consistent with this possible confound, we found that on average, higher-amplitude trajectories were rated as both more difficult and more enjoyable. This pattern raises the possibility that the observed difficulty-enjoyment association may partly reflect responses to movement dynamics rather than environmental constraints per se. Future work should disentangle these factors by exploring a wider range of trajectory shapes.

Despite this potential confound, viewer judgments were strikingly consistent given the brief exposure: participants viewed just 10 seconds of movement through individual obstacles, yet consistently extracted both difficulty and dangerousness. This rapid extraction likely depends on efficient perceptual processes: observers may track specific diagnostic cues such as obstacle spacing for difficulty or proximity to collision for dangerousness. The selective influence of difficulty on enjoyment also suggests that evaluation involves more than perception alone: viewers may weight perceived features by their relevance to evaluative goals, a form of goal-directed attention (Borji and Itti, 2014). Future work using eye-tracking methods could reveal whether participants spontaneously attend to different features when judging enjoyment versus difficulty.

More broadly, our approach illustrates the value of using procedurally generated stimuli to study action evaluation. By systematically manipulating environment difficulty and trajectory dangerousness within different movement patterns, we could test precise mechanistic hypotheses about what makes a performance enjoyable to watch. Our operationalizations — difficulty as the expected likelihood of success on a task, and dangerousness as the susceptibility to failure along a specific trajectory — could be extended to domains beyond simple arcade games, from board games to multi-agent settings (e.g., Collins et al., 2025).

Together, these results identify specific event-level features — environmental difficulty in particular — that predict spectators’ affective responses, and link the perception of action constraints to the affect they produce. Enjoyment of unfolding action is not ineffable: it tracks observable features of the events being watched.

## Acknowledgements

We thank Ian-Christopher Tanoh and members of the Cognitive Tools Lab at Stanford University for helpful discussion.

This work was supported by awards from the Brown Institute for Media and Innovation to J.P.C. and J.E.F.; NSF CAREER Award #2436199, NSF DRL #2400471 to J.E.F.; and awards from the Stanford Human-Centered AI Institute (HAI) and Stanford Accelerator for Learning to J.C. and J.E.F.

## Code availability

Code and materials are available at <https://github.com/cogtoolslab/enjoyable-action-sequences-cogsci26>.

## References

- Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349. <https://doi.org/10.1016/j.cognition.2009.07.005>
- Bar-Anan, Y., Wilson, T. D., & Gilbert, D. T. (2009). The feeling of uncertainty intensifies affective reactions. *Emotion*, *9*(1), 123–127. <https://doi.org/10.1037/a0014607>
- Barney, J., & Pennington, N. (2023). An exploration of esports fan identity, engagement practices, and motives. *Telematics and Informatics Reports*, *11*, 100081. <https://doi.org/10.1016/j.teler.2023.100081>
- Borji, A., & Itti, L. (2014). Defending yabus: Eye movements reveal observers' task. *Journal of Vision*, *14*(3), 29–29. <https://doi.org/10.1167/14.3.29>
- Collins, K. M., Zhang, C. E., Wong, L., Costa, M. B. d., Todd, G., Weller, A., Cheyette, S. J., Griffiths, T. L., & Tenenbaum, J. B. (2025, October). People use fast, flat goal-directed simulation to reason about novel problems [arXiv:2510.11503 [q-bio]]. <https://doi.org/10.48550/arXiv.2510.11503>
- Dennett, D. C. (1987). *The intentional stance* [Pages: xi, 388]. The MIT Press.
- Dredge, S. (2015). PewDiePie: How the YouTube king clocked up 40m fans and 10bn views. *The Guardian*. <https://www.theguardian.com/technology/2015/oct/22/pewdiepie-youtube-felix-kjellberg-this-book-loves-you>
- Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572–595). Oxford University Press.
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292. [https://doi.org/10.1016/S1364-6613\(03\)00128-1](https://doi.org/10.1016/S1364-6613(03)00128-1)
- Gjata, N. N., Ullman, T. D., Spelke, E. S., & Liu, S. (2022). What Could Go Wrong: Adults and Children Calibrate Predictions and Explanations of Others' Actions Based on Relative Reward and Danger. *Cognitive Science*, *46*(7), e13163. <https://doi.org/10.1111/cogs.13163>
- Google. (2014). Flappy bird, 2048, destiny. Retrieved January 27, 2026, from <https://archive.google/trends/2014/story/flappy-bird.html>
- Gudmundsson, S. F., Eisen, P., Poromaa, E., Nodet, A., Purmonen, S., Kozakowski, B., Meurling, R., & Cao, L. (2018). Human-Like Playtesting with Deep Learning [ISSN: 2325-4289]. *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, 1–8. <https://doi.org/10.1109/CIG.2018.8490442>
- Gweon, H., Asaba, M., & Bennett-Pierre, G. (2017). Reverse-engineering the process: Adults' and preschoolers' ability to infer the difficulty of novel tasks. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *39*(0).
- Hamilton, J. A., Haier, R. J., & Buchsbaum, M. S. (1984). Intrinsic enjoyment and boredom coping scales: Validation with personality, evoked potential and attention measures. *Personality and individual differences*, *5*(2), 183–193.
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*, 243–259. <https://doi.org/10.2307/1416950>
- Holdaway, C., Bear, D. M., Radwan, S. F., Frank, M. C., Yamins, D. L. K., & Fan, J. E. (2021). Measuring and predicting variation in the interestingness of physical structures. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43).
- Isaksen, A., Wallace, D., Finkelstein, A., & Nealen, A. (2017). Simulating strategy and dexterity for puzzle games. *2017 IEEE Conference on Computational Intelligence and Games (CIG)*, 142–149. <https://doi.org/10.1109/CIG.2017.8080427>
- Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The Naïve Utility Calculus: Computational Principles Underlying Commonsense Psychology. *Trends in Cognitive Sciences*, *20*(8), 589–604. <https://doi.org/10.1016/j.tics.2016.05.011>
- Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2020). The Naïve Utility Calculus as a unified, quantitative framework for action understanding. *Cognitive Psychology*, *123*, 101334. <https://doi.org/10.1016/j.cogpsych.2020.101334>
- Kaspar, K., Zimmermann, D., & Wilbers, A.-K. (2016). Thrilling News Revisited: The Role of Suspense for the Enjoyment of News Stories. *Frontiers in Psychology*, *7*, 1913. <https://doi.org/10.3389/fpsyg.2016.01913>
- Leonard, J. A., Bennett-Pierre, G., & Gweon, H. (2019). Who is better? Preschoolers infer relative competence based on efficiency of process and quality of outcome. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *41*(0).
- Liu, S., Karakose-Akbiyik, S., Outa, J., & Kim, M. J. (2026). How physical information is used to make sense of the psychological world. *Nature Reviews Psychology*, *5*(1), 59–73. <https://doi.org/10.1038/s44159-025-00514-1>
- Liu, S., Pepe, B., Ganesh Kumar, M., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2022). Dangerous Ground: One-Year-Old Infants are Sensitive to Peril in Other Agents' Action Plans. *Open Mind : Discoveries in Cognitive Science*, *6*, 211–231. [https://doi.org/10.1162/opmi\\_a\\_00063](https://doi.org/10.1162/opmi_a_00063)
- Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science*, *358*(6366), 1038–1041. <https://doi.org/10.1126/science.aag2132>
- Madrigal, R., Bee, C., Chen, J., & LaBarge, M. (2011). The Effect of Suspense on Enjoyment Following a Desirable Outcome: The Mediating Role of Relief. *Media Psychology*, *14*(3), 259–288. <https://doi.org/10.1080/15213269.2011.596469>
- Martinez, J., Binder, F. J., Wang, H., Haber, N., Fan, J. E., & Yamins, D. (2023). Measuring and Modeling Physical Intrinsic Motivation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45).
- Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and

- future development. *Emotion Review*, 5(2), 119–124. <https://doi.org/10.1177/1754073912468165>
- Moulard, J. G., Kroff, M., Pounders, K., & Ditt, C. (2019). The Role of Suspense in Gaming: Inducing Consumers' Game Enjoyment. *Journal of Interactive Advertising*, 19(3), 219–235. <https://doi.org/10.1080/15252019.2019.1689208>
- Parsons, S. (2021). Splithalf: Robust estimates of split half reliability. *Journal of Open Source Software*, 6(60), 3041. <https://doi.org/10.21105/joss.03041>
- Pirouz, D. M., Johnson, A. R., Thomson, M., & Pirouz, R. (2015). Creating Online Videos That Engage Viewers. *MIT Sloan Management Review*. Retrieved January 27, 2026, from <https://sloanreview.mit.edu/article/creating-online-videos-that-engage-viewers/>
- Pizzo, A. D., Baker, B. J., Na, S., Lee, M. A., Kim, D., & Funk, D. C. (2018). eSport vs. Sport: A Comparison of Spectator Motives [Publisher: SAGE Publications]. *Sport Marketing Quarterly*, 27(2), 108–123. <https://doi.org/10.32731/smq.272.062018.04>
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017, August). Proximal Policy Optimization Algorithms [arXiv:1707.06347 [cs]]. <https://doi.org/10.48550/arXiv.1707.06347>
- Smith, C. A., & Kirby, L. D. (2012). Affect and cognitive appraisal processes. In *Handbook of affect and social cognition* (pp. 76–93). Psychology Press.
- Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of behavioral decision making*, 15(4), 263–290.
- Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In *Blackwell handbook of childhood cognitive development* (pp. 167–187). Blackwell Publishing. <https://doi.org/10.1002/9780470996652.ch8>
- Yildirim, I., Saeed, B., Bennett-Pierre, G., Gerstenberg, T., Tenenbaum, J., & Gweon, H. (2019). Explaining intuitive difficulty judgments by modeling physical effort and risk. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 41(0).
- Zillmann, D., Bryant, J., & Sapolsky, B. S. (1989). Enjoyment from Sports Spectatorship. In *Sports, Games, and Play* (2nd ed.). Psychology Press.