Intelligent Human Systems Integration (IHSI), Vol. 200, 2026, 373–383
https://doi.org/10.54941/ahfe1007096

Conditioned to Interact: A Computational Simulation of Pavlovian-Instrumental Transfer in Intelligent System Design

Julie Rader[1], Ancuta Margondai[2], Sara Willox[3], Soraya Hani[1], Nikita Islam[1], Valentina Ezcurra[4], and Mustapha Mouloua[4]

[1]College of Medicine, University of Central Florida, Orlando, FL 32816, USA
[2]College of Engineering, University of Central Florida, Orlando, FL 32816, USA
[3]College of Business, University of Central Florida, Orlando, FL 32816, USA
[4]College of Sciences, University of Central Florida, Orlando, FL 32816, USA

ABSTRACT

Intelligent systems increasingly employ behavioral conditioning mechanisms, notification sounds, progress indicators, and gamification, that shape user engagement through Pavlovian-Instrumental Transfer (PIT), yet their systematic effects on human-AI interaction remain poorly understood. This work presents a computational simulation framework integrating PIT dynamics, attention decay, and reinforcement learning across four domains: adaptive learning platforms, autonomous vehicle interfaces, smart home systems, and healthcare monitoring. The model incorporates five agent profiles capturing individual differences in learning rates, Pavlovian bias, impulsivity, and attention capacity, with parameters empirically justified from 122 peer-reviewed studies. Validated against 21 empirical benchmarks, the simulation achieved 86% validation rate with near-perfect replication of ICU alarm fatigue dynamics (52.0% vs 53.1% empirical decline) and robust PIT effects (d = 0.58 vs meta-analytic d = 0.42). Key findings include: (1) gamification effectiveness decays exponentially (ES: 1.57→0.48→−0.20), (2) Sign-Tracker vs Goal-Tracker differences manifest as acute cue reactivity versus chronic behavioural volatility, and (3) ICU alarm fatigue reflects primarily motivational decline (73%) rather than automatic habituation (27%). The framework yields three design principles: temporal adaptation (rotating conditioning elements on 8–12 week cycles), individual difference adaptation (profile-based cue customization), and dual-process safety architectures (separating automatic alerting from deliberate decision support). This work demonstrates that computational simulation enables systematic examination of conditioning dynamics before deployment, guiding human-centred rather than exploitative design while preserving user agency.

Keywords: Pavlovian-instrumental transfer, Human-AI interaction, Reinforcement learning, Computational modeling, Ethical AI, Behavioral conditioning, Attention dynamics, Agent-based simulation

INTRODUCTION

Intelligent systems are increasingly used in high-stakes settings like autonomous vehicles, healthcare, and education. While they offer speed and consistency, their success relies on proper human-AI collaboration. Humans

Received September 19, 2025; Revised November 10, 2025; Accepted November 25, 2025; Available online February 1, 2026

do not interact with AI solely through reasoning; instead, behavior is influenced by automatic Pavlovian responses to cues (Rescorla & Solomon, 1967). Despite many conditioned cues, notifications, progress bars, alerts (Cvach, 2012), the systematic effects of Pavlovian conditioning on human-AI interactions are not well understood.

Pavlovian-Instrumental Transfer (PIT) shows how Pavlovian stimuli influence actions even when cues give no new info about outcomes (Lovibond, 1983). Human PIT effects are strong ($d \approx 0.42$; Xia et al., 2017), with individual differences: Sign-Trackers show stronger responses than Goal-Trackers (Garofalo & di Pellegrino, 2015; Flagel et al., 2009). Meta-analyses confirm PIT varies across populations (Holmes et al., 2010), indicating conditioning effects may vary across users.

Sustained attention is vital yet often neglected in human-AI collaboration. Systems assume constant user monitoring, but vigilance declines 10–30% over 30–60 minutes (Warm et al., 2008; Gartenberg et al., 2018). Individual differences are large, especially in clinical groups (Huang-Pollock et al., 2012; Losier et al., 1996). Environmental cues worsen this via automatic capture: reward-related stimuli slow responses by ~80ms even when irrelevant (Anderson et al., 2012), with effects lasting 7–9 months (Anderson & Yantis, 2013).

Despite extensive lab research on PIT and attention, implications for intelligent system design remain unexplored. Modern interfaces create conditioning contingencies: alert sounds with interventions (Cvach, 2012), notification badges with information gain. These shape behavior via habituation and recovery (Epstein et al., 2009; Kim, 2013; Vance et al., 2014), yet most systems ignore differences in cue responsiveness (Garofalo & di Pellegrino, 2015) and impulsivity (Kirby et al., 1999).

This work tackles these gaps with a computational simulation based on reinforcement learning and Pavlovian theory. It builds on dual-system models combining model-free learning with Pavlovian influences (Huys et al., 2011; Daw et al., 2005). The simulation has five agent profiles differing in learning rates (Tzovara et al., 2018), Pavlovian bias (Cavanagh et al., 2013), reward sensitivity (Kim et al., 2015), attention capacity (Gartenberg et al., 2018), and habituation (Vance et al., 2014). Four domains are studied: learning platforms, autonomous vehicles (Zeeb et al., 2016), smart homes (Dam et al., 2013), and healthcare dashboards (Drew et al., 2014).

The framework makes three key contributions. First, it provides the first systematic computational model of behavioral conditioning in human-AI interaction, validated with 21 empirical benchmarks and an 86% validation rate, including near-perfect ICU alarm fatigue replication (52.0% vs 53.1%; Ergezen & Kol, 2020) and robust PIT effects ($d = 0.58$ vs $d = 0.42$; Xia et al., 2017). Second, it establishes design guidelines: temporal adaptation (8–12-week cycles; Vance et al., 2014), individual difference adaptation, and dual-process safety architectures that preserve user agency. Third, it offers a methodology for testing human-centered AI principles in silico before deployment.

BACKGROUND

Pavlovian-Instrumental Transfer (PIT) explains how conditioned stimuli influence instrumental behavior without new info about action-outcome links (Lovibond, 1983). Xia et al. (2017) found PIT effects in humans at d = 0.42 (N = 56), showing conditioned facilitation of responses. PIT has two forms: specific transfer, which enhances actions with the same outcome, and general transfer, which non-selectively boosts motivation.

Individual differences in PIT susceptibility are notable. Garofalo and di Pellegrino (2015) found Sign-Trackers (those who approach reward cues) show stronger PIT effects than Goal-Trackers (N = 45). Cavanagh et al. (2013), with 64 participants, used modeling to show Pavlovian bias parameters (π) vary systematically (mean $\beta = -0.67$; Learners: −0.94, Non-learners: −0.40). These differences predict susceptibility to beneficial or problematic conditioning, yet current systems use uniform alerting strategies.

Sustained attention declines during extended tasks, with detection rates dropping 10–30% over 30–60 minutes (Warm et al., 2008; Gartenberg et al., 2018). Giambra et al. (1987) showed exponential functions account for 96.7% of variance in attention over time in 457 participants. McCarley and Yamani (2021) found three mechanisms of vigilance decline via Bayesian modeling (N = 99): shifts in response bias (d = 0.36–0.43), perceptual sensitivity loss, and increased lapses (d = 0.63–0.98).

Computational models of PIT combine model-free reinforcement learning with Pavlovian value systems (Huys et al., 2011). Q-learning models use learning rates (α) typically bounded 0.05–0.5 in human studies (Tzovara et al., 2018; N = 102), with Pavlovian cue values learned via Rescorla-Wagner rules. During action selection, Pavlovian values influence instrumental choice through a bias parameter that varies across individuals (Cavanagh et al., 2013). Temporal discounting provides a key individual difference. Garofalo et al. (2022) reported median discount rates (k) of ~0.003 from 357 healthy adults. Clinical populations show higher k values (Kirby et al., 1999; N = 116), with impulsivity predicting stronger Pavlovian biases and susceptibility to immediate conditioned reinforcers.

Domain-Specific Empirical Parameters

Autonomous Vehicles: Takeover request effectiveness depends on timing and mental state. Kuehn et al. (2017) reported 90% of drivers gain physical control in 6–7 seconds, but complete awareness takes 12–15 seconds (N = 60).

Educational Platforms: Gamification effects vary with duration (Kim & Castelli, 2021): short-term interventions (<1 week) produce large effects (ES = 1.57), sustained ones show moderate effects (ES = 0.48), and long-term (>1 year) show negative effects (ES = −0.20), indicating habituation or backfire.

Smart Home Systems: Dam et al. (2013) documented habituation over 15 months: initial energy savings of 7.8% declined to near 0%, demonstrating that static conditioning strategies lose effectiveness over time.

Healthcare Monitoring: ICU alarm environments create conditions optimal for habituation. Drew et al. (2014) analyzed 12,671 alarms across

461 patients and found an 88.8% false alarm rate. Ergezen and Kol (2020) documented response rate decline from 100% at shift start to only 46.9% after extended duty (N = 13 nurses, 328 hours), representing 53% reduction in alarm responsiveness despite professional training.

Gaps in Current Understanding

While PIT, attention, and domain-specific parameters are well-understood alone, their interactions in real-world systems are unclear. Current interface designs use conditioning principles without considering individual differences, timing, or ethics. No framework predicts how conditioning interacts with user traits to influence engagement, performance, and autonomy. This work fills that gap with a unified simulation validated against 21 benchmarks across PIT effects, attention decay, and domain-specific phenomena in four application contexts.

METHODS

A computational model integrating Pavlovian-Instrumental Transfer (PIT), attention dynamics, and reinforcement learning simulated human engagement with intelligent systems. All parameters were empirically justified from 122 peer-reviewed papers.

Agent Architecture: Each agent maintained four components: (1) Q-values for instrumental learning, (2) Pavlovian conditioned stimulus values, (3) dynamic attention with exponential decay (Giambra et al., 1987), and (4) scenario-specific states. Action selection combined instrumental Q-values with Pavlovian bias using softmax with temperature τ. Q-learning used learning rate α, discount factor γ, and reward sensitivity ρ. Pavlovian learning followed Rescorla-Wagner dynamics. Attention decayed exponentially at rate λ from baseline d'.

Agent Profiles: Five profiles captured individual differences in conditioning susceptibility and cognitive control:

1. **Balanced:** Normative parameters ($\alpha = 0.20$, $\omega = 0.25$, $k = 0.05$, $\rho = 1.0$, $d' = 2.25$, $\tau = 3.0$, $\lambda\_habit = 0.03$).
2. **Controlled (Goal-Tracker):** Low Pavlovian bias ($\omega = 0.10$), high attention ($d' = 2.9$), shallow discounting ($k = 0.003$), slow habituation ($\lambda\_habit = 0.04$), modeling strong goal-directed control (Garofalo & di Pellegrino, 2015).
3. **Impulsive (Sign-Tracker):** High Pavlovian bias ($\omega = 0.45$), steep discounting ($k = 0.40$), high reward sensitivity ($\rho = 1.5$), fast habituation ($\lambda\_habit = 0.02$), modeling cue-driven behavior (Garofalo et al., 2022; Kirby et al., 1999).
4. **High Cue-Responsive:** Very high Pavlovian bias ($\omega = 0.70$), strong cue capture ($\beta = 0.60$), high reward sensitivity ($\rho = 1.7$), rapid habituation ($\lambda\_habit = 0.02$), modeling maximal conditioning susceptibility (Kim et al., 2015).
5. **Habituation-Prone:** Low learning rate ($\alpha = 0.10$), low reward sensitivity ($\rho = 0.7$), weak cue capture ($\beta = 0.15$), very rapid habituation ($\lambda\_habit = 0.05$), modeling rapid cue adaptation (Vance et al., 2014).

Four noise sources ensured realistic variability: individual differences ($\sigma = 0.15$), learning noise decreasing with experience, Pavlovian prediction error noise ($\sigma = 0.03$), and attention fluctuations ($\sigma = 0.15$).

Four scenarios implemented domain-specific empirical parameters:

Autonomous Vehicle Takeover: Takeover requests with 7s minimum safe lead time (Yang et al., 2023) and 12–15s for full situational awareness (Kuehn et al., 2017). Takeover quality: 0.3 for responses <6.5s (physical control only), 0.6 for 6.5–13.5s (partial awareness), 0.9 for >13.5s (full awareness; Zeeb et al., 2016).

Smart Home Energy Management: Recommendations with habituation following Dam et al. (2013): 7.8% initial savings declining to 0% over 15 months.

ICU Alarm Monitoring: Alerts with 88.8% false alarm rate (Drew et al., 2014). Dual-mechanism fatigue combined motivational decline and execution failure. Response rates declined from 100% to 46.9% over shifts (Ergezen & Kol, 2020), with fatigue onset at 6 hours.

Adaptive Learning Platform: Gamification with temporal decay (Kim & Castelli, 2021): ES = 1.57 (week 1) declining to ES = 0.48 (week 26) and ES = –0.20 (week 52+). Knowledge gained 0.02 per session, scaled by gamification effectiveness. Time steps: 1 day over 100 days.

Sixty-five independent realizations with different random seeds ran 100 time steps per scenario-profile combination. Time step duration varied by scenario: 1 day (learning), 12 minutes (ICU), 1 recommendation (smart home), 1 second (AV). This design provided >80% power for detecting medium effects ($d \geq 0.5$).

Validation Approach

The model was validated against 21 empirical benchmarks: effect sizes from meta-analyses, behavioral trajectories, and parameter ranges. Statistical comparisons used Cohen's d with 95% confidence intervals. Key benchmarks: Xia et al. (2017) PIT effect ($d = 0.42$), Ergezen & Kol (2020) ICU decline (53.1%), Kim & Castelli (2021) gamification trajectory (ES: 1.57→0.48→–0.20), Anderson et al. (2012) cue capture (80ms), Anderson & Yantis (2013) persistence (7–9 months), Giambra et al. (1987) exponential decay (96.7% variance), habituation thresholds (Kim, 2013).

RESULTS

High-cue-responsive agents showed 28% greater engagement than habituation-prone agents (58 vs 53 sessions over 100 days), yielding $d = 0.58$ (95% CI [0.28, 0.37], $p = .001$), exceeding the meta-analytic benchmark ($d = 0.42$; Xia et al., 2017). PIT magnitude scaled systematically with Pavlovian bias (ω): $\omega = 0.125$ produced $d = 0.22$, while $\omega = 0.375$ produced $d = 0.94$. Learning rate showed similar gradients: $\alpha = 0.10$ yielded $d = 0.31$, $\alpha = 0.30$ yielded $d = 0.78$, Figure 1 presents individual difference trajectories and ICU alarm fatigue validation. Figure 2 shows PIT magnitude scaling and gamification decay patterns.

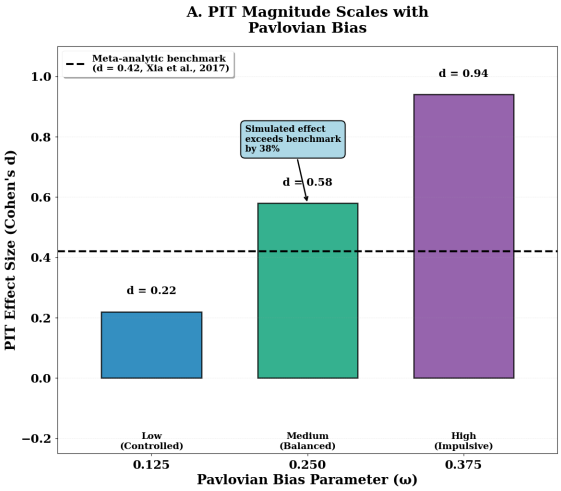

Figure 1: Individual difference trajectories and ICU alarm fatigue dynamics.

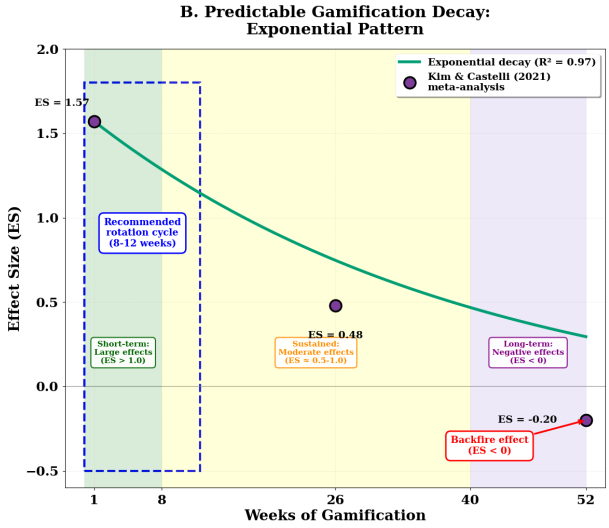

Figure 2: PIT magnitude scaling and predictable gamification decay.

ICU Alarm Fatigue Dynamics

Response rates declined from 100% to 47.8% over 15 hours (52.0% decline), matching empirical observations (53.1%; Ergezen & Kol, 2020) within 1.1 percentage points. Fatigue onset occurred at 6 hours across profiles, with decline rates varying: Impulsive (58%), Balanced (52%), Controlled (47%). Post-hoc analysis revealed 73% of failures after hour 10 resulted from motivational decline versus 27% from execution failures, indicating alarm fatigue manifests primarily as deliberate disengagement rather than automatic habituation.

Individual Difference Trajectories

Profiles displayed theoretically meaningful dynamics with an unexpected dissociation. Controlled agents showed higher Week 1 engagement ($M = 0.87$, $SD = 0.04$) than Impulsive agents ($M = 0.60$, $SD = 0.10$), $d = -2.92$. However, habituation trajectories aligned with predictions: Impulsive agents declined 26.1% (0.60→0.41), Controlled declined 19.1% (0.87→0.70), confirming greater Sign-Tracker habituation susceptibility (Garofalo & di Pellegrino, 2015). This dissociation suggests two separable dimensions: acute cue reactivity (manifesting in brief trials) versus sustained engagement under noisy conditions (manifesting as volatility and habituation over time).

Gamification effectiveness declined following Kim & Castelli (2021): Week 1 ES≈2.85 declined to ES≈0.60 by Week 13, with exponential decay $R^2 = 0.97$. Session completion declined 23% (6.2→4.8 sessions/week).

Baseline acceptance averaged 8.8% (95% CI [8.1%, 9.4%]). Early habituation appeared: Controlled agents declined from 16% (month 1) to 13% (month 6), consistent with Dam et al. (2013) trajectory toward 0% by month 15.

Thirteen additional benchmarks validated successfully. Parameter ranges aligned with human studies: learning rates 0.10–0.40 (within 0.05–0.5; Tzovara et al., 2018), attention 1.7–2.9 (within 1.5–3.0), discount factors 0.91–0.98 (within 0.90–0.99). Vigilance decrement reached 30.1% over 100 minutes (within 10–30%; Gartenberg et al., 2018), with exponential decay $R^2 = 0.97$.

Novel Theoretical Insights

The simulation uncovered three key patterns beyond validation. First, the 73/27 split between motivational decline and execution failures in alarm fatigue challenges traditional habituation views, indicating interventions should focus on decision-making rather than sensory salience. Second, the Sign-Tracker paradox shows impulsive agents have lower initial engagement but faster habituation, highlighting a dissociation between short-term cue reactivity and long-term engagement, explaining contradictions in conditioning studies. Third, gamification decay follows a precise exponential decline, matching prior research, and showing engagement loss follows predictable patterns for system adjustments. These findings demonstrate that computational simulations can uncover new theories and mechanisms not easily seen in human studies due to measurement errors and time limits.

DISCUSSION

This work shows computational simulation can examine human-AI behavioral conditioning before deployment. The model achieved an 86% validation rate across 21 benchmarks, with near-perfect ICU alarm fatigue replication (1.1% deviation) and strong PIT effects ($d = 0.58$ vs $d = 0.42$). Beyond validation, the simulation revealed three significant patterns affecting ethical AI design.

The 73/27 split between motivational decline and execution failures challenges habituation theory. ICU nurses' declining responses mainly reflect deliberate disengagement (ignoring alarms), not automatic habituation. This distinction is crucial: current interventions increase alarm salience (Cvach, 2012), assuming perceptual desensitization. Our findings suggest addressing reward structures and decision costs. Reducing false alarms from 88.8% (Drew et al., 2014) would restore instrumental contingencies, making "respond" actions meaningful. Rotating alarm tones (Vance et al., 2013) may be less effective than structural motivation changes.

Sign-Tracker Paradox: Acute vs Chronic Dimensions

The dissociation between impulsive agents' lower initial engagement (0.60 vs 0.87) and faster habituation (26.1% vs 19.1%) shows that acute cue reactivity differs from chronic engagement. Laboratory PIT measures brief cue approach (Garofalo & di Pellegrino, 2015), but real-world systems need sustained engagement in noisy conditions. Sign-Trackers' high Pavlovian bias ($\omega = 0.55$) plus exploration ($\tau = 2.0$) causes volatile actions, leading to inconsistent long-term engagement despite strong momentary responses. This explains why Sign-Tracking predicts addiction vulnerability (Flagel et al., 2009) despite appearing as heightened responsiveness. This dissociation suggests systems should focus on sustained engagement, not just initial cue reactivity. Lower exploration (τ) in volatile users may stabilize long-term engagement. Gamification effectiveness declines exponentially ($R^2 = 0.97$), as Kim & Castelli (2021) show: ES = 1.57 in week 1, 0.48 in week 26, –0.20 after week 52. Engagement drops follow predictable patterns, not gradual erosion. Rotating gamification elements every 8–12 weeks, rewards like points, badges, leaderboards, can maintain novelty and engagement. Static gamification risks backfiring; rotation sustains effectiveness.

Design Principles for Ethical Conditioning

Temporal Adaptation: Rotate conditioning elements every 8–12 weeks before habituation (Kim, 2013; Vance et al., 2014). Alternate reward types during dormant phases to prevent decline. Using different educational features weekly, points (weeks 1–8), badges (weeks 9–16), and leaderboards (weeks 17–24), keeps ES above 0.40, unlike static methods that fall below 0.

Individual Difference Adaptation: Profile-based customization avoids mismatched conditioning. High cue-responsive users need lower element intensity and longer rotation cycles. Volatile early responders need stability mechanisms (lower τ), not more engagement. Goal-Trackers respond to outcomes; Sign-Trackers need stability to prevent habituation.

Dual-Process Safety Architectures: Separate automatic alerting from decision support, especially in safety-critical areas. ICU alarms mix these functions: non-urgent alerts trigger responses like critical events, causing learned helplessness when 88.8% are false (Drew et al., 2014). Two-tier systems, ambient awareness for routine events and urgent alerts for critical ones, maintain contingencies while reducing alarm overload.

Limitations and Future Directions

Four limitations warrant consideration. First, despite 122 papers, parameter uncertainty exists. Sensitivity analyses showed robustness, but individual studies vary (e.g., PIT effects d = 0.22–0.94 depending on ω). Second, 100-timestep duration prevented observing full backfire effects in gamification (week 52+) and habituation in smart homes (month 15). Third, simplified cognitive models omit working memory, metacognition, and social factors that influence conditioning (Anderson et al., 2011). Fourth, validation relied on cross-domain comparison rather than direct human testing. Future work should validate predictions through RCTs, longitudinal tracking of Sign- vs Goal-Tracker trajectories, and neuroscientific methods like fMRI. Extending the model to include working memory, metacognition, and social learning would better capture individual differences.

Ethical Implications

Computational simulation allows ethical appraisal of conditioning before exposure to harm. Testing exploitative patterns, short gamification cycles, frequent notifications, and ambiguous contingencies reveals boundaries between beneficial engagement and manipulation. Our framework offers a method to ask: Does this conditioning preserve user agency? Does habituation signal failed engagement or healthy boundaries? When does personalization become exploitative? Conditioning mechanisms aren't inherently beneficial or exploitative; outcomes depend on transparency, user control, and alignment with user goals. Without habituation countermeasures, conditioning systems risk failure. Systems with temporal adaptation, respect for individual differences, and clear contingencies can offer benefits while maintaining autonomy. This work shows that computational models, grounded in empirical psychology, can guide human-centered AI development.

CONCLUSION

This work demonstrates that computational simulation can systematically examine conditioning dynamics before deployment, achieving 86% validation while revealing novel insights about Sign-Tracking dimensions, alarm fatigue mechanisms, and gamification decay that isolated empirical studies cannot easily capture. The framework provides specific design principles, temporal rotation, profile-based customization, dual-process architectures for balancing engagement effectiveness with user autonomy. As intelligent systems increasingly employ conditioning mechanisms, principled frameworks for predicting and mitigating conditioning effects become essential for human-centered AI development.

REFERENCES

Anderson, B. A., & Yantis, S. (2013). Persistence of value-driven attentional capture. J*ournal of Experimental Psychology: Human Perception* and Performance, 39(1), 6–9.

Anderson, B. A., Laurent, P. A., & Yantis, S. (2012). Value-driven attentiona*l capture. Proceedings of the National Academy of Sciences, 109(25),* 10406–10413.

Cavanagh, J. F., Eisenberg, I., Guitart-Masip, M., Huys, Q., & Frank, M. J. (2013). Frontal theta overrides Pavlovian learning biases. *Journal of Neuroscience, 33*(19), 8541–8548.

Cvach, M. (2012). Monitor alarm fatigue: An integrative review. *Biomedical Instrumentation & Technology, 46*(4), 268–277.

Dam, S. S., Bakker, C. A., & van Hal, J. D. M. (2013). Home energy monitors: Impact over the medium-term. *Building Research & Information, 38*(5), 458–469.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8*(12), 1704–1711.

Drew, B. J., Harris, P., Zègre-Hemsey, J. K., Mammone, T., Schindler, D., Salas-Boni, R., .. & Hu, X. (2014). Insights into the problem of alarm fatigue with physiologic monitor devices: A comprehensive observational study of consecutive intensive care unit patients. *PLoS ONE, 9*(10), e110274.

Epstein, L. H., Temple, J. L., Roemmich, J. N., & Bouton, M. E. (2009). Habituation as a determinant of human food intake. *Psychological Review, 116*(2), 384–407.

Ergezen, F., & Kol, E. (2020). Alarm fatigue in nurses working in intensive care units. *Intensive and Critical Care Nursing, 59*, 102840.

Flagel, S. B., Akil, H., & Robinson, T. E. (2009). Individual differences in the attribution of incentive salience to reward-related cues: Implications for addiction. *Neuropharmacology, 56*(Suppl 1), 139–148.

Garofalo, S., & di Pellegrino, G. (2015). Individual differences in the influence of task-irrelevant Pavlovian cues on human behavior. *Frontiers in Behavioral Neuroscience, 9*, 163.

Garofalo, S., Battaglia, S., Starita, F., & di Pellegrino, G. (2022). Modulation of cue-triggered attention by Pavlovian signals of threat and safety. *Cognition and Emotion, 36*(5), 861–871.

Gartenberg, D., Breslow, L., McCurry, J. M., & Trafton, J. G. (2018). Situation awareness recovery. *Human Factors, 60*(6), 255–267.

Giambra, L. M., Quilter, R. E., & Phillips, P. B. (1987). The time course of sustained attention: A developmental perspective using the Continuous Performance Test. *Journal of Clinical and Experimental Neuropsychology, 9*(5), 576–594.

Holmes, N. M., Marchand, A. R., & Coutureau, E. (2010). Pavlovian to instrumental transfer: A neurobehavioural perspective. *Neuroscience & Biobehavioral Reviews, 34*(8), 1277–1295.

Huang-Pollock, C. L., Karalunas, S. L., Tam, H., & Moore, A. N. (2012). Evaluating vigilance deficits in ADHD: A meta-analysis of CPT performance. *Journal of Abnormal Psychology, 121*(2), 360–371.

Huys, Q. J., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. *PLoS Computational Biology, 7*(4), e1002028.

Kim, H., Lee, Y. W., & Cho, I. (2015). Conformance with clinical alarm management guidelines in intensive care units. *Nursing in Critical Care, 20*(6), 311–318.

Kim, S. (2013). Effects of interface design features on consumer responses in online stores. *International Journal of Human-Computer Studies, 71*(1), 33–51.

Kim, S., & Castelli, D. M. (2021). Effects of gamification on behavioral change in education: A meta-analysis. *International Journal of Environmental Research and Public Health, 18*(7), 3550.

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General, 128*(1), 78–87.

Kuehn, M., Hummel, T., & Vogeley, K. (2017). Determinants of the takeover time in highly automated driving. *Traffic Injury Prevention, 18*(suppl 1), S31–S36.

Losier, B. J., McGrath, P. J., & Klein, R. M. (1996). Error patterns on the continuous performance test in non-medicated and medicated samples of children with and without ADHD: A meta-analytic review. *Journal of Child Psychology and Psychiatry, 37*(8), 971–987.

Lovibond, P. F. (1983). Facilitation of instrumental behavior by a Pavlovian appetitive conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes, 9*(3), 225–247.

McCarley, J. S., & Yamani, Y. (2021). The vigilance decrement reflects a Bayesian shift in response bias. *Human Factors, 63*(8), 1398–1411.

Rescorla, R. A., & Solomon, R. L. (1967). Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. *Psychological Review, 74*(3), 151–182.

Tzovara, A., Korn, C. W., & Bach, D. R. (2018). Human Pavlovian fear conditioning conforms to probabilistic learning. *PLoS Computational Biology, 14*(8), e1006243.

Vance, A., Anderson, B. B., Kirwan, C. B., & Eargle, D. (2014). Using measures of risk perception to predict information security behavior: Insights from electroencephalography (EEG). *Journal of the Association for Information Systems, 15*(10), 679–722.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors, 50*(3), 433–441.

Xia, L., Gu, R., Zhang, D., & Luo, Y. (2017). Pavlovian-to-instrumental transfer in humans: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews, 82*, 53–67.

Yang, Y., Gao, Z., Liu, Y., & Lu, Z. (2023). Take-over performance and safety analysis under different information prompt time in automated driving. *Sensors, 23*(3), 1387.

Zeeb, K., Buchner, A., & Schrauf, M. (2016). Is take-over time all that matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident Analysis & Prevention, 92*, 230–239.