

Effects of Enemy Visual-Behavior Consistency on Players' Expected Experience in High-Tension Levels: A Scenario-Based Pilot Study

Shih-Chieh, Liao^{1*}, Peng-Xu, Pan²

Department of Multimedia and Computer Entertainment Science, Southern Taiwan University of Science and Technology, Tainan, Taiwan

Corresponding Email: ¹alfietw@stust.edu.tw, ²asahipan216@gmail.com

Abstract

In high-tension action levels, players often rely on enemy visual cues, such as body silhouette, implied movement speed, and attack wind-up cues, to anticipate threats. However, designers may also intentionally create contrast between enemy appearance and attack behavior in order to produce surprise, pressure, or stronger combat tension. This pilot study examined how enemy visual-behavior consistency influenced players' expected experience in high-tension levels, and evaluated whether a scenario-based image questionnaire could serve as a low-cost early screening method for enemy concepts. Eight enemy stimuli were developed, including four high-consistency and four low-consistency designs. Each stimulus paired a static enemy visual with a short high-tension combat vignette, and four questionnaire versions were used to control presentation order. Nineteen valid responses were collected with no missing data. Participants rated perceived consistency, expected tension/pressure, and overall expected experience for each stimulus. Results showed that high-consistency designs produced higher perceived consistency than low-consistency designs ($M = 5.695$ vs. $M = 4.428$, Wilcoxon signed-rank test, $p = .0004$), indicating that the manipulation was recognizable. Low-consistency designs produced higher expected tension/pressure than high-consistency designs ($M = 5.047$ vs. $M = 4.408$, $p = .0066$). More notably, low-consistency designs also produced higher overall expected experience than high-consistency designs ($M = 5.086$ vs. $M = 4.605$, $p = .0175$). These findings suggest that high consistency mainly supports perceived reasonableness, predictability, and readability, whereas low consistency can amplify tension, stimulation, and overall expected experience in high-tension contexts. Low consistency should therefore not be treated only as a design flaw; when used with an appropriate context, it may function as a tension-amplification strategy.

Keywords: expected player experience, enemy design, visual-behavior consistency, high-tension levels, scenario-based evaluation

1. Introduction

In action games, enemy encounters are often interpreted before players fully experience their behavior. Players make rapid judgments from silhouette, apparent weight, implied speed, and attack wind-up cues. These judgments influence whether an incoming threat is perceived as predictable, exciting, dangerous, or unfair. From an affordance perspective, gameplay depends on the player's ability to perceive possible actions and understand how a game object can or should be engaged. In digital games, this interpretation involves the virtual representation, the player's prior experience, and embodied expectations about possible action outcomes (Rambusch & Susi, 2008). When enemy visual cues and behavior are aligned, players can form stable expectations. When visual cues and behavior contradict each other, expectation is disrupted; this disruption may increase pressure, but it may also become stimulation, surprise, or challenge anticipation.

This study focuses on enemy design in high-tension levels. In this paper, high-tension levels are operationalized as combat contexts with strong threat salience, such as high damage risk, costly mistakes, limited resources, or time pressure. In such contexts, players are already primed to monitor danger and predict enemy actions quickly. Therefore, whether enemy cues are readable and whether they fit the assigned behavior can directly affect how players interpret an encounter. Gameplay experience research has consistently indicated that challenge, flow, and immersion are closely related but difficult to separate conceptually (Ermi & Mäyrä, 2005; Michailidis et al., 2018; Nacke & Lindley, 2008). In enemy design, visual readability may help players understand a threat and focus on adaptation and mastery, whereas contrast between appearance and behavior may increase uncertainty and perceived tension even before direct interaction occurs.

However, enemy design does not always require maximum consistency. From a practical design perspective, high-consistency enemies are useful for establishing rules, reducing interpretation cost, and supporting quick player judgment. Low-consistency enemies may be used for special enemies, elite enemies, mini-bosses, or high-tension segments in order to break existing expectations and create a stronger sense of danger or memorability. Enemy visual-behavior consistency should therefore be treated not simply as a matter of correct or incorrect design, but as a design variable that shapes players' expected experience.

Testing enemy readability only after a fully playable prototype has been implemented can be expensive. Once models, animations, effects, values, and encounter scripts are integrated, revising the core concept of an enemy may require substantial production resources. Early-stage game development therefore needs low-cost methods that allow designers to screen promising or risky concepts before full implementation. Prior work on paper prototyping and prototype fidelity shows that early evaluation can provide meaningful experience-related feedback before major resources are committed (Roque, 2010; Sim et al., 2013). Building on this logic, the present study proposes a scenario-based image questionnaire that uses static enemy visuals and short high-tension combat descriptions to evaluate players' expected responses to enemy consistency designs.

Accordingly, this study asks two research questions. First, how does enemy visual-behavior consistency influence perceived consistency, expected tension/pressure, and overall expected experience in a high-tension combat context? Second, can a scenario-based image questionnaire serve as a low-cost screening tool for enemy concepts during early production, especially for identifying whether a contrast design may function as a tension amplifier?

2. Literature Review

2.1 Immersion, challenge, and expected gameplay experience

Immersion is one of the most frequently discussed concepts in game studies, but it has been defined in multiple ways. Ermi and Mäyrä (2005) proposed the SCI model, which distinguishes sensory immersion, challenge-based immersion, and imaginative immersion. This framework is especially relevant to the present study because enemy readability is closely related to challenge-based immersion. When the demands of the game are legible enough for players to invest attention, adapt, and gradually master the encounter, players are more likely to become absorbed in a challenge-oriented state. Nacke and Lindley (2008) also argued that experiential states such as flow, immersion, excitement, and challenge need clear operationalization rather than loose descriptive use.

Later reviews have emphasized that flow and immersion are often treated as overlapping constructs, especially in studies focusing on concentrated attention, task absorption, and the balance between skill and demand (Michailidis et al., 2018). In practical design terms, a combat encounter does not need to be easy in order to feel positive, but it generally needs to appear interpretable enough for players to believe that they can form a response strategy. Challenge is therefore not only a function of raw difficulty; it is also shaped by how demands are presented and perceived. Denisova et al. (2020) further showed that challenge is a multidimensional player experience that may involve cognitive, emotional, performative, and decision-making components. For enemy design, this means that contrast between visual cues and behavior may increase emotional or cognitive load before direct interaction takes place.

Because the present study uses non-playable image-text stimuli, it examines expected gameplay experience rather than post-play experience. This distinction is important. Expected tension, expected pressure, and overall expected experience are anticipatory judgments formed before actual gameplay. Although they cannot fully represent actual play, they remain valuable in early design evaluation. When an enemy concept has not yet entered full production, players' initial judgments about whether it seems reasonable, threatening, or worth encountering can help designers decide whether the concept should be retained, revised, or discarded.

2.2 Visual cues, affordances, and enemy design consistency

Enemy interpretation begins with cues. Body scale, apparent mass, implied speed, and attack preparation all function as signals that help players estimate danger and predict behavior. Rambusch and Susi (2008) argued that game affordances are not simply located in on-screen objects themselves; rather, they are perceived through the interaction among virtual representation, learned convention, and embodied player understanding. This is directly relevant to enemy design, because players use prior game experience and visual conventions to judge whether a large enemy is likely to be slow and powerful, whether a small enemy is likely to be fast and evasive, or whether a visible wind-up suggests a predictable heavy attack.

Research on game visuals further suggests that graphical presentation influences player experience rather than merely decorating gameplay. Gerling et al. (2013) found that graphical fidelity can change player impressions and may matter more strongly when mechanics are demanding. Related work has also linked visual style to perceived challenge, indicating that players evaluate challenge not only through mechanics, but also through the visual language that frames interaction. Although this literature does not examine enemy visual-behavior consistency directly, it supports the broader argument that visual presentation shapes expectations about quality, competence, and difficulty.

From a design standpoint, consistency can be understood as the degree to which an enemy's visual language fits its assigned behavior. A high-consistency enemy communicates an internally coherent signal: a heavy-looking enemy performs forceful and high-damage attacks, or an agile-looking enemy performs fast movement and evasive behavior. A low-consistency enemy intentionally breaks this expectation, such as when a large and heavy-looking enemy suddenly performs a fast dash, or when a small and agile-looking enemy performs an unexpectedly heavy attack. Such contrast design may create surprise, threat escalation, and memorability, but it may also be interpreted as unfair or confusing. Low consistency is therefore not automatically a design flaw; its effect depends on level context, frequency of use, and whether players can interpret the contrast as a meaningful challenge.

2.3 Early-stage evaluation and low-fidelity prototype logic

A persistent problem in game production is that many experience-related problems are detected too late. Once enemy models, animations, effects, and encounter scripts have been integrated into a playable level, revising a weak concept can become expensive. Roque (2010) argued that paper prototyping and rehearsal can be used to evaluate gameplay experience in early design stages before substantial resources are committed. The value of such methods lies not in perfectly reproducing the final game, but in revealing whether a concept is clear, understandable, and directionally promising.

Prototype fidelity research also suggests that lower-fidelity materials can still produce useful evaluative feedback, although different fidelities may foreground different kinds of responses. Sim et al. (2013) showed that evaluation outcomes can vary across prototype fidelities, which means early methods must be interpreted according to their purpose. When the

research question concerns expectation, conceptual fit, and directional preference rather than fine-grained control, reaction time, or motor performance, low-fidelity methods can still be useful.

The scenario-based image questionnaire proposed in this study follows this logic. It does not attempt to replace playable testing. Instead, it functions as an early screening tool for design risk and conceptual direction. If players interpret an enemy concept as reasonable, stimulating, and worth encountering at the expectation level, the concept may deserve further implementation. If players interpret a design as confusing, overly oppressive, or unfair, designers can adjust it before formal production. In this way, scenario-based evaluation can reduce trial-and-error costs at the front end of production while helping teams clarify the design intention of enemy concepts.

3. Method

3.1 Research design

This study employed a scenario-based pilot design in which participants evaluated eight enemy concept stimuli within the same high-tension combat context. Each stimulus consisted of a static enemy visual and a short textual description of a high-risk level situation. The central manipulation was the consistency between enemy visual cues and assigned attack behavior. The visual cue dimensions included body silhouette, implied movement speed, and attack wind-up cues. Among the eight stimuli, four were designed as high-consistency stimuli and four as low-consistency stimuli.

Participants were not asked to judge visual quality or art preference in isolation. Instead, they were asked to imagine confronting each enemy in the described high-tension combat situation and then rate their expected gameplay experience. The method therefore captured anticipatory design judgments before direct gameplay.

To reduce possible order effects, four questionnaire versions were used. Version A presented the stimuli in the following order: H1, L1, H2, L2, H3, L3, H4, L4. Version B used L2, H2, L3, H3, L4, H4, L1, H1. Version C used H3, L3, H4, L4, H1, L1, H2, L2. Version D used L4, H4, L1, H1, L2, H2, L3, H3. This rotation was intended to reduce the influence of a fixed sequence on participants' responses.

3.2 Participants

Nineteen valid questionnaire responses were collected. Versions A, B, and C each received five responses, and Version D received four responses. No missing values were found. Participants had basic game experience and were able to understand enemy encounter situations in side-scrolling or action game contexts. Because the study was designed as a pilot, the sample was used mainly to examine directional patterns and methodological feasibility rather than to establish a full inferential model. Participation was voluntary and anonymous.

3.3 Stimulus preparation and validation

The stimulus construction process began from a design matrix that combined enemy visual cues and attack behaviors. High-consistency stimuli paired appearance and behavior in an intuitively matched way, such as a large and heavy-looking enemy paired with oppressive or high-damage behavior, or an agile-looking enemy paired with fast movement and evasive behavior. Low-consistency stimuli deliberately assigned behaviors that contradicted visual implications of size, speed, or attack preparation, such as a large enemy that suddenly performed a fast dash, or an agile-looking enemy that delivered a heavy attack.

Before the questionnaire was administered, the eight materials were screened through expert review to verify two issues. First, each image-text pair needed to be recognizable as a high-tension combat situation. Second, the intended difference between high-consistency and low-consistency conditions needed to be sufficiently clear. This validation step was important because the usefulness of a scenario-based image questionnaire depends on whether players can recognize the design differences embedded in the stimuli.

3.4 Measures and procedure

After viewing each stimulus, participants rated three main dimensions on seven-point scales: perceived consistency, expected tension/pressure, and overall expected experience. Perceived consistency was used to examine whether participants felt that the enemy's visual cues and behavior were mutually coherent. Expected tension/pressure evaluated whether the enemy seemed threatening, oppressive, or uncertain in the high-tension context. Overall expected experience assessed whether participants considered the encounter attractive and worth playing.

Because the study used non-playable static visuals and textual descriptions, the questionnaire wording emphasized expected experience rather than retrospective gameplay experience. The selected constructs were informed by prior work on immersion, challenge, and questionnaire-based player experience assessment (Denisova et al., 2020; Ermi & Mäyrä, 2005; Poels et al., 2007).

The procedure was straightforward. Participants first read a shared high-tension scenario that established the high-risk combat context. They then viewed the eight stimuli in the order assigned by their questionnaire version and completed the rating items for each stimulus. Finally, the stimuli were regrouped into high-consistency and low-consistency categories, and mean scores were calculated for each construct.

3.5 Data analysis

Descriptive statistics and Wilcoxon signed-rank tests were used for analysis. Because the sample size was 19 and the study was positioned as a pilot study, the analysis focused on overall trends and design implications between high-consistency and low-consistency conditions rather than a complete multifactorial model. The Wilcoxon signed-rank test was appropriate for comparing two related conditions rated by the same participants and provided a conservative preliminary examination for small-sample data.

4. Results

Nineteen valid responses were analyzed. Versions A, B, and C each had five responses, and Version D had four responses. No missing data were identified. After data cleaning, the eight stimuli were regrouped into high-consistency and low-consistency conditions. Descriptive statistics and Wilcoxon signed-rank tests were then conducted. Because the study remained a pilot study, the results were interpreted as directional evidence and design implications rather than as a final confirmatory model.

For perceived consistency, high-consistency designs produced a mean score of 5.695, whereas low-consistency designs produced a mean score of 4.428. The difference was statistically significant, $p = .0004$. This result indicates that participants were able to recognize the intended difference between high-consistency and low-consistency enemies. It also suggests that the manipulation of enemy visual-behavior consistency had preliminary validity. When enemy body scale, speed implication, and attack behavior were mutually coherent, players were more likely to interpret the enemy as reasonable and predictable.

For expected tension/pressure, low-consistency designs produced a higher mean score than high-consistency designs. The low-consistency mean was 5.047, whereas the high-consistency mean was 4.408. The difference was statistically significant, $p = .0066$. This result indicates that when enemy appearance and behavior contradicted each other, players expected stronger tension and pressure. In a high-tension level context, low consistency was not immediately interpreted as a design error. Instead, unpredictability and contrast appeared to amplify players' sense of danger.

The most notable result appeared in overall expected experience. Low-consistency designs produced a mean score of 5.086, whereas high-consistency designs produced a mean score of 4.605. This difference was also statistically significant, $p = .0175$. This result does not fully support the intuitive assumption that high consistency always produces a better overall experience. Although high consistency provided clearer and more reasonable enemy logic, the uncertainty, stimulation, and pressure created by low-consistency designs may have made these enemies more attractive in the high-tension context.

Taken together, the results suggest that the effect of enemy design consistency is not one-directional. High-consistency designs are suitable for establishing basic readability and combat grammar. Low-consistency designs may function as tension amplifiers in high-tension levels. In other words, consistency primarily affects whether players regard an enemy as reasonable, clear, and predictable, whereas low consistency may affect whether players regard an enemy as dangerous, stimulating, and worth challenging.

5. Discussion

The results indicate that enemy visual-behavior consistency is not a simple linear determinant of better or worse experience. High-consistency designs scored significantly higher in perceived consistency than low-consistency designs, showing that players could use body silhouette, implied speed, and attack wind-up cues to judge whether enemy behavior seemed reasonable. This supports the basic assumption of the study: when enemy appearance and attack behavior align, players can form more stable expectations and understand the enemy's combat role more easily.

The more distinctive finding is that low-consistency designs increased not only expected tension/pressure, but also overall expected experience. This suggests that players in high-tension levels do not necessarily prefer enemies that are fully reasonable and fully predictable. When enemy visual cues and behavior create contrast, players may experience stronger uncertainty and threat. In a high-tension context, this feeling may be transformed into stimulation, surprise, and challenge anticipation. In this study, low consistency was therefore not simply interpreted as confusion or failure. Instead, it demonstrated a usable design value.

This finding has practical implications for enemy design. If all enemies in a game use high-consistency design, players may understand the rules easily, but level pacing may become relatively stable and surprise may decrease. By contrast, carefully introduced low-consistency design may increase player alertness, tension, and challenge anticipation when applied to elite enemies, mini-bosses, ambush enemies, or high-tension segments. However, low consistency still requires control. If contrast appears too frequently or without contextual framing, players may lose trust in the combat language of the game and interpret uncertainty as unfairness.

Therefore, this study suggests treating high consistency and low consistency as two different design tools. High consistency is appropriate for establishing the basic grammar of an enemy system, such as tutorial encounters, ordinary enemies, and combat situations that require quick readability. Low consistency is more suitable for intensifying emotional peaks in a level, such as special enemies, sudden enemy transformation, high-tension scenes, or pre-boss escalation. The key design question is whether players at that moment need clear understanding or renewed alertness. If the design goal is to teach rules and support comprehension, high consistency is more suitable. If the design goal is to increase pressure and stimulation, low consistency may have design value.

The study also demonstrates the usefulness of a scenario-based image questionnaire for early screening. Before a team produces enemy models, animations, and combat scripts, designers can use static visuals and short scenarios to test whether players understand the enemy's intended logic and whether a contrast design creates the expected tension. This method cannot replace playable prototype testing, but it can reduce early development cost and help designers decide which enemy concepts deserve further production.

The results also point to a need to distinguish productive tension from unfair stress. Low-consistency designs improved overall expected experience in the current study, suggesting that players may interpret them as interesting challenges. However, this does not mean that all contrast designs will produce positive effects. If low consistency lacks visual hints, level context, or learning space, it may still cause confusion and frustration. Future studies should therefore combine playable data, error rates, reaction time, and interviews to identify when low consistency creates tension and when it damages perceived fairness.

6. Limitations and Future Work

Several limitations should be acknowledged. First, the study used a scenario-based image questionnaire, so it measured expected experience formed after viewing static enemy visuals and textual descriptions rather than actual play experience. The findings therefore reflect players' first impressions and anticipatory judgments, but they cannot fully capture operational pressure, reaction time, failure learning, animation readability, or recovery behavior during real combat.

Second, the sample size was 19. This is appropriate for a pilot study and directional interpretation, but it is insufficient for a more complete multifactorial validation model. Although the current results show significant differences between high-consistency and low-consistency designs in perceived consistency, expected tension/pressure, and overall expected experience, future studies should recruit a larger sample to examine interactions among player skill, game experience, and enemy type.

Third, the current study compared high-consistency and low-consistency as two broad categories. It did not separately model the effects of body silhouette, implied speed, and attack behavior. Future research could decompose enemy design into body size, movement speed, and behavior factors in order to identify which cue most strongly influences tension, readability, or overall experience.

Fourth, the present analysis used overall mean comparisons and did not conduct a full test of questionnaire version effects. Although four versions were used to control stimulus order, the small sample size limited further examination of whether version differences influenced responses. Future studies with larger samples should include version as a control variable to improve the stability of the results.

Future research can extend the present work in three directions. First, the eight enemy stimuli should be implemented in a playable prototype, such as a Unity action game prototype, so that expected experience can be compared with actual gameplay experience. Second, player ability indicators, such as reaction time, play history, self-rated skill, or level performance, should be included to examine whether high-skill and low-skill players interpret low-consistency designs differently. Third, behavioral or physiological measures, such as error frequency, dodge delay, heart rate, or electrodermal activity, could be used to distinguish effective tension from excessive stress.

7. Conclusion

Enemy appearance is not merely a visual layer placed over mechanics. It is an important interface through which players understand threat, predict behavior, and form combat expectations. This study examined the effect of enemy visual-behavior consistency on players' expected experience in high-tension levels. Results showed that high-consistency designs effectively increased perceived consistency, making enemies appear more reasonable, recognizable, and understandable. Low-consistency designs increased expected tension/pressure and, in the present study, also increased overall expected experience.

These findings indicate that low consistency does not necessarily equal design failure in high-tension levels. When placed in an appropriate context, contrast between appearance and behavior may become a strategy for increasing stimulation, threat, and enemy memorability. High consistency is suitable for building rules and maintaining readability, whereas low consistency is suitable for creating sudden shifts, pressure, and surprise within high-tension segments.

The contribution of this study is both conceptual and methodological. Conceptually, it frames enemy visual-behavior consistency as a design variable that affects expected player experience. Methodologically, it demonstrates the potential of a scenario-based image questionnaire for early-stage game production. With this method, designers can examine whether an enemy concept has a clear design intention and whether a contrast design is likely to be interpreted as productive tension before investing in models, animations, and full combat systems. Future work combining playable prototypes, player ability grouping, and behavioral or physiological data can further build a more complete model of enemy consistency and high-tension level experience.

Table 1. Operationalization of Enemy Design Consistency

Cue dimension	High-consistency example	Low-consistency contrast example
Body silhouette	Large and heavy-looking enemy paired with forceful, high-damage behavior; agile-looking enemy paired with agile behavior.	Large or heavy-looking enemy paired with unexpectedly fast or evasive behavior; small or agile-looking enemy paired with unexpectedly heavy attacks.
Implied movement speed	Visual posture and overall shape imply a movement profile that fits the assigned combat role.	Visual posture suggests one mobility profile, but the assigned behavior implies another.
Attack wind-up cue	Preparatory pose communicates a readable attack timing and direction.	Preparatory cue is ambiguous, delayed, or contradictory relative to the eventual attack pattern.

Table 2. Descriptive Summary of Current Questionnaire Results

Construct	High-consistency result	Low-consistency result	Statistical result	Design interpretation
Perceived consistency	M = 5.695	M = 4.428	p = .0004	High-consistency designs were more likely to be perceived as reasonable, clear, and recognizable.
Expected tension/pressure	M = 4.408	M = 5.047	p = .0066	Low-consistency designs increased uncertainty and threat, making them suitable for high-tension contexts.
Overall expected experience	M = 4.605	M = 5.086	p = .0175	Low consistency did not reduce overall experience in this pilot study; instead, it increased expected appeal in high-tension enemy encounters.

Note. N = 19. The table reports pilot-level descriptive patterns. The p-values are based on Wilcoxon signed-rank tests and should be interpreted as preliminary evidence rather than a final validation model.

References

- Denisova, A., Cairns, P., Guckelsberger, C., & Zendle, D. (2020). Measuring perceived challenge in digital games: Development & validation of the challenge originating from recent gameplay interaction scale (CORGIS). *International Journal of Human-Computer Studies*, 137, 102383. <https://doi.org/10.1016/j.ijhcs.2019.102383>
- Ermi, L., & Mäyrä, F. (2005). Fundamental components of the gameplay experience: Analysing immersion. In J. Jenson (Ed.), *Changing Views: Worlds in Play - Selected Papers of the 2005 Digital Games Research Association's Second International Conference* (pp. 15-27). Digital Games Research Association & Simon Fraser University.
- Gerling, K. M., Birk, M., Mandryk, R. L., & Doucette, A. (2013). The effects of graphical fidelity on player experience. In *Proceedings of the International Conference on Making Sense of Converging Media* (pp. 229-236). ACM. <https://doi.org/10.1145/2523429.2523473>
- Michailidis, L., Balaguer-Ballester, E., & He, X. (2018). Flow and immersion in video games: The aftermath of a conceptual challenge. *Frontiers in Psychology*, 9, 1682. <https://doi.org/10.3389/fpsyg.2018.01682>
- Nacke, L. E., & Lindley, C. A. (2008). Flow and immersion in first-person shooters: Measuring the player's gameplay experience. In *Proceedings of the 2008 Conference on Future Play* (pp. 81-88). ACM. <https://doi.org/10.1145/1496984.1496998>
- Poels, K., de Kort, Y. A. W., & IJsselstein, W. A. (2007). D3.3: Game Experience Questionnaire: Development of a self-report measure to assess the psychological impact of digital games. Eindhoven University of Technology.
- Rambusch, J., & Susi, T. (2008). The challenge of managing affordances in computer game play. *Human IT*, 9(3), 83-109.
- Roque, L. G. (2010). Early game design rehearsal with paper prototyping. In *IX Simpósio Brasileiro de Games e Entretenimento Digital (SBGames 2010)*.
- Sim, G., Cassidy, B., & Read, J. C. (2013). Understanding the fidelity effect when evaluating games with children. In *Proceedings of the 12th International Conference on Interaction Design and Children* (pp. 193-200). ACM. <https://doi.org/10.1145/2485760.2485769>