COLOMBO SCIENTIFIC PUBLISHING

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE & INFORMATION SYSTEM
E-ISSN: 2536-7919
P-ISSN: 2536-7900

DOI: - https://doi.org/10.55640/ijcsis/Volume11Issue03-01 **PAGE NO: 01-07**

Using Virtual Game Spaces as Experimental Grounds for Testing Hypotheses

Mikhail Petrov

Game designer, Founder & CEO NEW EDGE DWC LLC
Dubai, UAE

ARTICLE INFO

Article history:

Submission: January 03, 2026
Accepted: February 17, 2026
Published: March 12, 2026
VOLUME: Vol.11 Issue 03 2026

Keywords:
virtual game spaces, game design as research, parameterized environments, experimental hypothesis testing, agent-based modeling, multi-agent learning, emergence, behavioral patterns, research games, complex systems

ABSTRACT

The article examines virtual game spaces as research testbeds for evaluating hypotheses about the behavior of complex systems. Its relevance follows from the growing number of domains in which field observation is constrained by ethics, logistics, or limited parameter control, while modeling still demands empirical calibration. The contribution lies in an analytical account of game design as a research practice aligned with systems science: the designer constructs a parameterized environment, specifies rules, agent types, feedback channels, and observable metrics, then interprets resulting regimes as data suitable for refining assumptions. The paper outlines principles for staging "game experiments," describes procedures for linking observed patterns to distributional models and dynamic processes, and explains how emergent effects arise and how they can be translated into testable statements. To achieve this aim, the study applies analysis of recent scholarship, comparison of methodological approaches, and conceptual modeling of the whole research cycle. The material is grounded in scientific publications from 2022 to –2025. The article targets researchers who use simulations, agent-based models, and game-derived methods in social, economic, and social-ecological problem settings.

Introduction

The expanding agenda of complex-systems analysis increasingly confronts a methodological gap: field observation often produces incomplete evidence, whereas laboratory protocols frequently lack the interaction density and environmental richness that generate system-level patterns. Virtual game spaces address this gap by combining controllable rules with high-frequency interactions. They grant researchers direct control over parameters and enable systematic recording of agent trajectories under variable outcomes, supporting empirical interrogation of mechanisms that remain opaque in natural settings. The article aims to explain how virtual game spaces are used to pose, test, and refine research hypotheses about complex systems when the designer adopts an investigator's stance rather than a content author.

The study pursues three objectives:

1) to describe the methodological cycle through which a parameterized game environment is constructed as an experimental setup;

2) to present ways of connecting observed in-game behavioral patterns to hypothesis statements and to models of system dynamics;

3) to analyze emergent effects as a source of empirical material and as a criterion for revising initial assumptions.

The novelty of the work stems from transferring the inferential logic of systems science into game design practice: rules and interfaces are treated as operationalizations of variables, while balancing is framed as one special case within a broader research-oriented design strategy focused on measurement, identifiability, and regime comparison.

Materials and Methods

The study relies on recent, peer-reviewed, and widely used research artifacts that treat games and game-like environments as controlled yet rich experimental systems. Adams, White, and Araujo [1] provide an example of using an online game to test opinion-dynamics predictions empirically. Agapiou et al. [2] present a multi-agent social-dilemma environment ("Melting Pot 2.0") that formalizes rule-driven interaction regimes suitable for hypothesis-driven evaluation. Freese and Lukosch [3] conceptualize adaptive game design for complex systems via a structured derivation pipeline ("Funnel of Game Design"), while Freese and Bekebrede [4] extend this logic by positioning game research by design as a mixed-methods research approach in project and system settings. Hertz et al. [5] propose an experimental approach grounded in multi-agent reinforcement learning for studying cognitive agents in social-ecological systems. Jeong and Lee [6] introduce an LLM-infused game system designed to observe emergent narrative phenomena within constrained design scaffolds. Laborie et al. [7] analyze role-playing game co-design as a way to surface "surprises" that carry diagnostic value for socio-ecosystem transitions. Motlagh, Horcea-Milcu, and König [8] map how serious games can be integrated into transformative sustainability research processes. Xu and Dang [9] examine emergent behaviors in a pursuit-evasion grid-world under MARL, offering a concrete taxonomy of coordinated actions. Yiannakoulias, Grignon, and Marshall [10] demonstrate how an online decision game can parameterize an agent-based model when real-world data are unavailable or non-generalizable.
Methods applied: comparative analysis of methodological claims across sources, structured synthesis of design-to-inference workflows, and conceptual modeling of an end-to-end research cycle (hypothesis → environment design → instrumentation → observation → inference → redesign).

Results

Virtual game spaces can be treated as experimental grounds when their design is framed as the construction of a parameterized system rather than a scripted experience. Across the analyzed literature, a convergent pattern emerges: the "game" functions as an executable specification of assumptions about agents, constraints, incentives, and information flow, while play traces act as observations produced by that specification. This reframing shift attention from content authoring to variable control, measurement design, and falsifiable expectations about system dynamics [3], [4]. In practice, the experimental value of a game space depends on whether the designer can map theoretical constructs to rule parameters, instrument the environment to capture relevant state transitions, and interpret resulting regimes as evidence for or against the hypothesis under study [10].

A recurring result is that hypothesis testing in game spaces rarely proceeds through a single "clean" manipulation; instead, it unfolds through iterative tightening of the environment's degrees of freedom. Freese and Lukosch [3] describe adaptive derivation from a messy problem situation toward a game model that makes a selected slice of the system observable. Freese and Bekebrede [4] develop this logic by treating game research by design as a research setup in which the artifact itself becomes an instrument that mediates between theoretical expectations and empirical material. In such a setup, a hypothesis is operationalized by specifying how outcomes should shift when parameters change (reward weights, interaction topology, resource regeneration, information visibility, action friction), and by declaring what constitutes a meaningful deviation from expectation (phase change, coordination breakdown, new equilibrium class).

A second result concerns the role of data when real-world measurement is missing, ethically constrained, or structurally non-transferable. Yiannakoulias, Grignon, and Marshall [10] show that an online decision game can elicit behavioral preferences that are then transformed into a decision model and transferred into

an agent-based simulation. The conceptual move is crucial: rather than treating play as "engagement," the study treats play as a structured elicitation protocol that yields parameters for agents' decision rules. This supports a broader methodological claim: game spaces can serve as controlled elicitation layers that bridge qualitative intent ("how agents choose") and quantitative simulation ("what flows result from those choices") [10]. Parallel logic appears in Adams, White, and Araujo [1], where an online game is designed so that observed opinion shifts can be compared with predictions from an opinion-dynamics model; the game serves as a measurement device tuned to variables that are hard to isolate in uncontrolled social interaction.

A third result centers on emergence as the primary epistemic output of such experimental grounds. In complex systems, the most informative behaviors often arise from interactions rather than isolated mechanics. Xu and Dang [9] provide a clear illustration: in a bounded grid-world pursuit-evasion setting, coordinated patterns can be decomposed into fundamental actions and their composites, revealing how multi-agent learning produces structured coordination rather than random adaptation. Coordination mode A and coordination mode B refer to two analytically distinct, pre-defined classes of multi-agent coordination patterns (identified through trace-based classification of joint action sequences), where each mode corresponds to a different stable interaction strategy that emerges under the same task but under different parameter settings (e.g., observability, reward coupling, or agent heterogeneity). The implication for hypothesis testing is that emergent coordination can be treated as a dependent variable: hypotheses can target the conditions under which coordination mode A (e.g., flanking-like behavior) appears more frequently than mode B, or under which coordination collapses into local greedy pursuit. Because these patterns are generated by repeated interaction under explicit constraints, they can be probed by altering observability, reward shaping, or agent heterogeneity and then testing whether the predicted shift in coordination regimes occurs [9], [2].

A fourth result is that the designer's investigative posture becomes explicit when games are built as "research laboratories" for socio-ecological and socio-technical dynamics. Hertz et al. [5] argue for experimental approaches that better preserve the characteristics of social-ecological systems by using multi-agent reinforcement learning to model interactions between cognitive agents and environmental structures. Within this framing, the environment is not background scenery; it is a coupled subsystem that shapes and is shaped by collective action. Hypotheses in such settings often concern how spatial/temporal affordances, resource feedback loops, or informational constraints drive collective outcomes (cooperation, exploitation, oscillations, lock-in). A practical result follows: virtual game spaces allow repeated trials over controlled parameter grids, supporting sensitivity analysis that would be infeasible in field contexts, while still producing rich interaction traces [5].

A fifth result concerns co-design and treating "surprises" as data rather than noise. Laborie et al. [7] analyze role-playing game co-design with scientists and stakeholders, emphasizing that unexpected outcomes during play sessions can reveal hidden assumptions, mismatched mental models, or unaccounted constraints in socio-ecosystem transition problems. From a hypothesis-testing standpoint, such surprises function as diagnostic anomalies: they point to missing variables, wrong causal arrows, or overly narrow operationalization. This aligns with the broader idea that, in complex systems, model validity is often improved not by polishing balance but by exposing the system to conditions that make latent structure visible [7], [3].

A sixth result concerns the integration of generative systems and the observation of emergent narrative or social dynamics under partially open-ended interaction. Jeong and Lee [6] propose an LLM-infused game system designed for emergent narrative, where gameplay can be observed as the system attempts to maintain coherence amid player-driven variability. For hypothesis testing, such environments enable questions about how scaffolding constraints (prompt structure, rule enforcement, memory bounds) influence the distribution of narrative outcomes and failure modes. The research value stems from making "design knobs" explicit and measurable, allowing designers to test claims such as: tighter scaffolds reduce fragmentation but suppress novelty, while looser scaffolds increase novelty but raise inconsistency rates [6].

Figure 1 integrates the above results into a single research workflow that treats game design as an experimental setup, aligning adaptive design logic [3], game research by design [4], and game-to-model

parameter transfer [10].

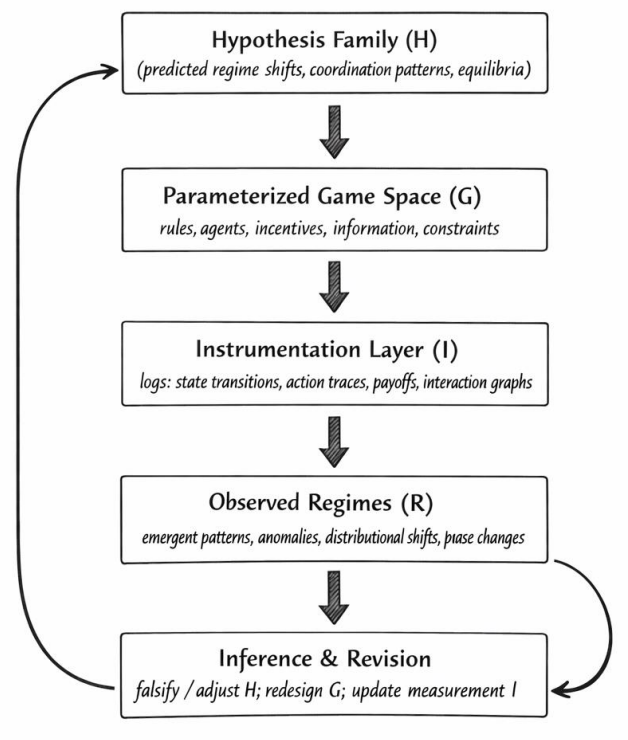

Figure 1. Research cycle for hypothesis testing in virtual game spaces (conceptual synthesis based on [3], [4], [10])

Finally, a synthesis emerges concerning how "balance" differs from a research-driven experimental posture. In balancing, parameter tuning targets user experience stability; in research-driven design, parameter tuning targets inferential leverage: maximizing the sensitivity of observable outputs to the variables that encode the hypothesis. This leverage is achieved by controlling confounds (e.g., reducing degenerate dominant strategies), ensuring identifiability (distinct parameter settings yield distinct regime signatures), and preserving enough richness for emergence without losing interpretability [3], [4], [2]. Motlagh, Horcea-Milcu, and König [8] complement this by positioning serious games inside broader research processes, clarifying that the experimental ground is not merely the game artifact but the coupled sequence of design, facilitation, observation, and interpretation steps that translate play into knowledge claims.

Discussion

The results support a methodological positioning of virtual game spaces as experimental grounds in which designers operate as investigators of complex systems, working with model families, distributions, and coupled dynamics rather than with predetermined scripts. A central implication is that validity hinges on the traceability from hypothesis constructs to rule parameters and from measurements to interpretable regime signatures. Evidence across sources indicates that this traceability strengthens when the environment is built around explicit variables (observability, incentive gradients, interaction topology, environmental feedback) and when the data pipeline is planned as part of design rather than appended after prototyping [3], [4], [10].

Table 1 consolidates how different classes of hypotheses map to design "knobs" and measurable outputs. The mapping is compiled from repeated methodological themes in adaptive game design for complex systems [3], parameter transfer from research games to ABM [10], multi-agent social dilemma benchmarking [2], and emergent MARL behavior analysis [9].

Table 1. Hypothesis classes, operationalization in game-space design, and measurable outputs (compiled from [2], [3], [9], [10])

Hypothesis class	Operational variable(s) in the game space	Observable output(s)	Typical test logic
Coordination emergence	Reward coupling; partial observability; agent heterogeneity	Coordination mode frequencies; time-to-coordination; stability of joint policies	Predict regime shift when observability or reward coupling crosses a threshold [9]
Decision-rule plausibility	Choice architecture; constraints on actions; information availability	Fitted decision model parameters; predictive accuracy on held-out traces	Compare the inferred decision rule with the theoretical expectation; transfer to simulation [10]
Social dilemma stability	Resource dynamics; punishment/sanction channels; payoff structure	Cooperation rates, oscillations, and collapse into defection	Test whether equilibrium class changes under altered sanctions or resource regeneration [2]
Robustness to perturbations	Noise; stochastic events; changes in topology	Resilience metrics; recovery time; variance of outcomes	Evaluate sensitivity surfaces over parameter sweeps [3]
Policy scaffolding effects	Rule enforcement strength; narrative constraints; memory limits	Failure taxonomy; coherence metrics; novelty distribution	Test predicted trade-off between coherence and novelty under scaffolding adjustments [6]

A methodological tension becomes visible: richer environments promote emergence, yet interpretability requires constraint. The reviewed approaches address this tension by iteratively narrowing the design space, treating each iteration as a sharper "lens" on the hypothesized mechanism [3], [4]. In this sense, emergence is not chased for spectacle; it is cultivated as a diagnostic signal whose structure can be compared across controlled changes.

A second implication concerns the status of stakeholder-driven role-play and co-design. Laborie et al. [7] show that surprises during role-playing sessions can expose hidden assumptions and mismatched system representations. This is compatible with the analytical reading of surprises as anomaly signals: when observed play violates expectations, the hypothesis family is likely underspecified, or the operationalization misses a relevant variable. Serious-game integration work in sustainability research further supports treating the entire process—design, facilitation, data capture, and interpretation—as the research setup, rather than merely the software artifact [8]. This expands what counts as "experimental control": control can be executed through facilitation protocols and scenario framing, not only through code.

Table 2 synthesizes threats to inference that arise specifically in virtual game spaces, alongside mitigation strategies grounded in the reviewed literature, including structured adaptive design [3], research-by-design framing [4], ABM parameterization via games [10], and MARL-based experimental modeling [5].

Table 2. Inference threats in virtual game spaces and mitigation strategies (derived from [3], [4], [5], [10])

Threat to inference	How it manifests in game-space studies	Mitigation strategy
Construct drift	Rules capture a different construct than the hypothesis intended	Explicit construct-to-parameter mapping; redesign until identifiability improves
Confounded outcomes	Multiple mechanics jointly drive the same outcome signature	Factorial design over parameters; ablation of mechanics; sensitivity surfaces
Overfitting to artifact	Conclusions hold only for a specific rule set or interface	Use hypothesis families; test robustness under controlled perturbations
Measurement blind spots	Logs miss critical state transitions or interaction context	Instrumentation planned as a design component; traceability requirements
Interpretability loss under richness	Emergent behavior appears but cannot be explained	Constrain degrees of freedom; iterative narrowing; regime labeling schemes

The distinction between balancing and research-driven design becomes sharper. Balancing optimizes playability metrics; research-driven design optimizes inferential contrast by seeking configurations in which outcome distributions respond diagnostically to parameter shifts that encode theoretical claims. This reframing aligns with the use of benchmark-like environments for social interaction regimes, where the goal is not entertainment stability but the controlled comparison of emergent collective behaviors under systematic variation. In LLM-infused systems, the same logic extends to narrative emergence: constraints become experimental factors, and narrative outcomes become distributions to compare rather than storylines to author.

Conclusion

Virtual game spaces function as experimental grounds when they are treated as parameterized systems with explicit measurement pipelines, enabling hypothesis testing through controlled variation and interpretable regime signatures. The analysis shows that the experimental cycle is strengthened by adaptive derivation from problem to game model and by treating the artifact as an instrument; research games can elicit behavioral parameters transferable to agent-based simulations, supporting hypotheses about decision processes in settings where real-world data are absent or non-generalizable; emergence serves as empirical material when it is operationalized into measurable patterns and compared across parameter shifts, as illustrated by multi-agent coordination regimes and social dilemma environments; co-design and stakeholder play generate anomaly signals (“surprises”) that guide hypothesis revision and variable expansion, while serious-game integration frameworks clarify how games embed into broader research processes; MARL-based experimental approaches and LLM-infused systems extend the space of testable hypotheses toward coupled social-ecological dynamics and emergent narrative behavior under constrained scaffolding.

References

1. Adams, J. A., White, G., & Araujo, R. P. (2022). Person-to-person opinion dynamics: An empirical study using an online game. PLOS ONE, 17(10), e0275473. https://doi.org/10.1371/journal.pone.0275473

2. Agapiou, J. P., Madhushani, U., Sunehag, P., Duéñez-Guzmán, E. A., Köster, R., Haas, J., … Leibo, J. Z. (2022). Melting Pot 2.0 (arXiv:2211.13746). arXiv.

3. Freese, M., & Lukosch, H. K. (2024). The Funnel of Game Design: An adaptive game design approach for complex systems. Simulation & Gaming, 55(2), 323–341. https://doi.org/10.1177/10468781231222524

4. Freese, M., & Bekebrede, G. (2025). Game research by design in project management and beyond: Conceptualization and methodology for serious gaming research set-ups. Project Leadership and Society, 6, 100174. https://doi.org/10.1016/j.plas.2024.100174

5. Hertz, U., Köster, R., Janssen, M., & Leibo, J. Z. (2025). Beyond the matrix: Experimental approaches to studying cognitive agents in social-ecological systems. Cognition, 254, 105993.

6. Jeong, J., & Lee, T. Y. (2025). LIGS: Developing an LLM-infused game system for emergent narrative. In Extended Abstracts of the 2025 CHI Conference on Human Factors in Computing Systems (CHI EA '25) (Paper 369). Association for Computing Machinery. https://doi.org/10.1145/3706599.3720212

7. Laborie, N., Flores, M. M., Thorin, M., Daré, W., & Bommel, P. (2025). Tackling surprises when co-designing by playing role-playing games: Application of the Sumak Kawsay game for understanding socio-ecosystem transitions in the Bolivian Altiplano. Ecology and Society, 30(3), 21. https://doi.org/10.5751/ES-16132-300321

8. Motlagh, M., Horcea-Milcu, I. A., & König, B. (2025). Discovering the potential of serious games for transformative sustainability research. Discover Sustainability, 6, 1–? https://doi.org/10.1007/s43621-024-00756-8

9. Xu, S., & Dang, Z. (2025). Emergent behaviors in multiagent pursuit evasion games within a bounded 2D grid world. Sci Rep. 15(1):29376. doi: 10.1038/s41598-025-15057-x. PMID: 40789927; PMCID: PMC12340135.

10. Yiannakoulias, N., Grignon, M., & Marshall, T. (2024). Parameterizing agent-based models using an online game. Computers, Environment and Urban Systems, 112, 102142. https://doi.org/10.1016/j.compenvurbsys.2024.102142