

Simulating Cognitive Bias in Interactive Environments: A Framework for Using Behavioural Simulation Games as Educational Tools.

Abstract:

Cognitive biases, systematic patterns of deviation from rational judgment, are well documented in psychological research, yet teaching people to recognise and resist them remains a genuinely unsolved problem. Traditional instruction, lectures explaining a bias and giving textbook examples, has been shown repeatedly to produce weak and short-lived improvement, in part because bias operates automatically and unconsciously, in exactly the moments a learner is not consciously applying what they were taught. This paper proposes a framework for designing interactive simulation games that let a user experience a cognitive bias operating on their own decisions in real time, rather than only reading a description of it, and argues that this experiential approach addresses a specific weakness in conventional bias education. The paper reviews the psychological foundation for this approach, proposes a concrete design framework mapping specific game mechanics to specific biases, discusses the technical architecture required to build and evaluate such a system, and examines the genuine limitations and ethical questions this kind of simulation raises. The paper concludes that interactive simulation is not a guaranteed cure for cognitive bias, but it is a substantially better-matched tool for the problem than passive instruction, and outlines what a rigorous implementation would actually require.

Keywords

cognitive bias, behavioural simulation, serious games, experiential learning, game-based education, decision science, human-computer interaction

1. Introduction

A cognitive bias is a systematic, predictable pattern in how people deviate from purely rational judgment, first systematically documented in the research programme Daniel Kahneman and Amos Tversky began in the 1970s. Decades of subsequent research have confirmed that biases such as anchoring, confirmation bias, loss aversion, and overconfidence are not occasional errors made by careless thinkers, but consistent features of normal human cognition, present even in experts and even in people who have been explicitly taught about the specific bias in question.

This last point is the central problem this paper addresses. A substantial body of research has found that simply teaching someone about a bias, explaining anchoring, for example, and showing a textbook demonstration of it, produces only a modest and often short-lived reduction in that person's own susceptibility to it. Trained economists who can correctly define anchoring on an exam still show measurable anchoring effects in their own subsequent purchasing decisions. This gap between conceptual knowledge and behavioural resistance suggests that the format of conventional bias education, a passive explanation delivered before or after the biased judgment occurs, is poorly matched to a phenomenon that operates automatically and largely outside conscious awareness at the actual moment of decision.

This paper proposes that interactive simulation games are a more appropriately matched educational format, because they can place a learner's own decision-making, with its own biases intact, at the centre of the experience, and reveal the bias's influence immediately afterward, while the decision and its context are still fresh. Section two reviews the

psychological and educational-theory foundation for this claim. Section three proposes a concrete design framework connecting specific biases to specific game mechanics. Section four discusses the technical architecture required to build and measure such a system. Section five examines genuine limitations and open ethical questions.

2. Theoretical Foundation

2.1 Why Passive Instruction Underperforms

The weak transfer from bias knowledge to bias resistance is well explained by dual-process theories of cognition, which distinguish between a fast, automatic, intuitive mode of judgment and a slower, effortful, deliberate mode. Most cognitive biases are products of the fast system operating exactly as it is designed to, producing quick judgments using mental shortcuts that are usually useful and occasionally systematically wrong. Conventional bias instruction engages the slow, deliberate system, the system a learner uses while reading a textbook or listening to a lecture, but the bias itself operates in the fast system, at a different moment and under different cognitive conditions than the ones present during instruction. Teaching the slow system about a pattern that occurs in the fast system does not reliably reach into the fast system's actual operation.

2.2 Experiential Learning Theory

David Kolb's experiential learning theory offers a useful framework for why simulation might close this gap. Kolb describes learning as a cycle moving through concrete experience, reflective observation, abstract conceptualisation, and active experimentation, arguing that durable learning requires movement through the full cycle rather than abstract conceptualisation, the lecture-and-textbook mode, alone. A well-designed bias simulation can provide the concrete experience stage directly: the learner makes a real decision under realistic pressure, experiences its consequence, and only then moves into reflective observation and abstract conceptualisation with a genuine personal experience to reflect on, rather than a hypothetical example belonging to someone else.

2.3 Game-Based Learning and Serious Games

The broader field of game-based learning, and specifically the subfield of serious games, games designed with an explicit educational or training purpose rather than pure entertainment, provides additional grounding. Researcher James Paul Gee has argued that well-designed games are effective learning environments specifically because they situate knowledge inside meaningful action and immediate consequence, precisely the structure this paper argues cognitive bias education requires. Gamification researchers including Sebastian Deterding have further distinguished genuine game-based learning, which embeds the learning objective inside real gameplay mechanics, from superficial gamification, which merely adds points and badges to conventional instruction without changing its underlying passive structure. This distinction matters directly for the framework proposed here: a bias simulation succeeds only if the bias is embedded in the actual mechanic the player is using to make decisions, not merely decorated around a conventional lesson.

3. A Design Framework: Mapping Biases to Mechanics

The central design challenge is translating an abstract cognitive bias into a concrete game mechanic that causes a player to experience the bias operating on their own choices, without the mechanic feeling like an obvious, telegraphed trick. This section proposes specific mappings for several well-studied biases.

3.1 Anchoring Bias

Anchoring, the tendency for an initial number to disproportionately influence a subsequent judgment even when the initial number is arbitrary, can be simulated through a resource-estimation or negotiation mechanic in which the game privately randomises a visible 'reference number' shown to the player before they submit an estimate or offer of their own. Comparing player estimates across sessions that received different randomised reference numbers reveals the anchoring effect directly in the player's own recorded decisions, which can then be shown back to them.

3.2 Loss Aversion and the Sunk Cost Fallacy

Loss aversion, the tendency to weigh a loss roughly twice as heavily as an equivalent gain, and the closely related sunk cost fallacy, the tendency to continue investing in a failing course of action because of resources already committed, can be simulated through a resource-management mechanic in which the player has invested effort or in-game currency into a project that the simulation reveals, partway through, to have a declining probability of success. Tracking whether and when the player abandons the failing investment, relative to a rational stopping-point benchmark the simulation calculates internally, directly measures susceptibility to both biases in the same mechanic.

3.3 Confirmation Bias

Confirmation bias, the tendency to seek and favour information that confirms an existing belief while discounting disconfirming information, can be simulated through an investigation or diagnosis mechanic in which the player forms an early hypothesis and then chooses which of several available pieces of evidence to examine further, some confirming and some disconfirming their stated hypothesis. Logging which evidence type the player disproportionately selects reveals confirmation bias directly through their own information-seeking behaviour, rather than through a self-report question about it.

3.4 Overconfidence

Overconfidence, the tendency to be more certain of one's own judgment than accuracy actually warrants, can be simulated by asking players to provide a confidence interval, rather than a single point estimate, for a series of factual or in-game predictions, and then comparing the proportion of correct answers that actually fall inside each player's stated confidence interval against the proportion that should fall inside it if their confidence were well calibrated.

4. Technical Architecture Requirements

Beyond game design, delivering this framework credibly requires specific technical infrastructure, and this section outlines the core components a genuine implementation would need.

4.1 Decision Telemetry

Every decision point relevant to a targeted bias must be logged with sufficient context to later analyse it: the options presented, the option chosen, response time, and any relevant randomised or player-specific variables such as the anchoring reference number described above. This telemetry layer is the foundation the entire measurement approach depends on, and it must be designed into the game's core data model from the start, not added afterward, since retrofitting decision logging onto a system that was not built to capture it is far more error-prone than designing it in from the outset.

4.2 A Bias-Detection and Feedback Layer

A separate analysis layer must process the raw telemetry to compute each player's measured susceptibility to each targeted bias, for example, the correlation between the randomised anchor value and the player's own estimate, and translate this into feedback the player actually receives, ideally immediately after the relevant decision, while its context is still available in working memory, consistent with the experiential learning argument in section two. This feedback layer is what actually closes Kolb's learning cycle; without it, the simulation is simply a game that happens to involve biased decisions, not an educational tool that reveals them to the learner.

4.3 Adaptive Scenario Generation

A more advanced implementation could use a player's measured bias profile to adaptively select which future scenarios to present, for example, presenting additional anchoring-relevant decisions to a player whose data suggests strong anchoring susceptibility, in order to give that player more opportunity to notice and practise resisting the specific bias they are most vulnerable to, rather than presenting an identical, undifferentiated sequence of scenarios to every player regardless of their individual profile.

5. Measuring Educational Effectiveness, and Its Limits

A rigorous evaluation of this framework requires comparing a player's bias susceptibility before and after extended use of the simulation, ideally using a control group that receives only conventional instruction on the same biases, to isolate the specific contribution of the interactive, experiential format from the contribution of simply discussing the biases at all.

A genuinely important caution belongs here. As discussed in section two, even direct personal experience of a bias, including experience gained through a well-designed simulation, is not guaranteed to produce durable resistance to that bias in entirely new, unrelated future contexts. The same research showing that trained economists remain vulnerable to anchoring in their own lives should temper any claim that a simulation, however well designed, fully inoculates a learner against a bias going forward. The more defensible claim is narrower: a well-designed simulation very likely produces stronger and more durable awareness than passive instruction alone, not that it produces complete, permanent resistance. Evaluation design and any claims made to users should reflect this more limited, more honest claim rather than overselling the intervention's effect.

6. Ethical Considerations

A framework explicitly designed to trigger cognitive biases in players, in order to teach them about those biases, sits close to a framework that could instead be used to exploit those same biases for manipulation rather than education, and this proximity deserves direct acknowledgment rather than avoidance. The ethical distinction rests substantially on transparency and intent: an educational implementation should disclose to players, at minimum by the end of a session, which biases were being tested and how their own data was used, and should never use the same techniques to nudge players toward a commercial or persuasive outcome the platform itself benefits from. A design framework capable of teaching bias is, by construction, also capable of exploiting it, and any real implementation carries a genuine obligation to build in that transparency deliberately rather than treat it as optional.

7. Conclusion

Conventional cognitive bias education faces a structural mismatch: it engages a learner's deliberate, reflective thinking to teach about a phenomenon that actually operates in their fast, automatic thinking, at a different moment and under different cognitive conditions than instruction occurs. Interactive simulation games offer a more appropriately matched format, capable of placing a learner's own biased decision-making at the centre of the experience and revealing its influence immediately afterward, closing the experiential learning cycle in a way passive instruction cannot. This paper has proposed concrete mappings from specific biases to specific game mechanics, outlined the technical architecture a credible implementation requires, and argued that both the evaluation of such a system and any claims made about its effectiveness should remain honest about the real, evidence-based limits of any bias intervention, however well designed. The strongest defensible claim for this framework is not that it eliminates bias, but that it teaches it better than the alternative most people currently receive.

References

- Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. *Proceedings of the 15th International Academic MindTrek Conference*, 9–15.
- Gee, J. P. (2003). *What Video Games Have to Teach Us About Learning and Literacy*. Palgrave Macmillan.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
- Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291.
- Kolb, D. A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Prentice-Hall.
- Odean, T. (1998). Are investors reluctant to realize their losses? *Journal of Finance*, 53(5), 1775–1798.
- Thaler, R. H. (2015). *Misbehaving: The Making of Behavioral Economics*. W. W. Norton & Company.
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.