

Countering Hindsight Bias and Memory Decay During Debriefing Through Dual-track Simulation Game Design

Ajay Gupta

Skills Café, Bangalore, India

ajay@skills.cafe

Abstract: The effectiveness of simulation-based learning is influenced by the quality of debriefing and post-simulation reflection. Most debriefs are conducted at the end of a game-based learning exercise, with a facilitator asking learners guided questions to reflect on their actions during gameplay, the rationale for their decisions, and the outcomes achieved. Since the authentic reasons behind decisions are not faithfully captured during gameplay, the learner's account of those reasons is affected by hindsight bias and memory decay. This study examines whether documenting the decision rationale (Concurrent Decision Logging, or CDL) before outcome disclosure in a simulation-based game alters participants' reasoning and subsequent explanations of their decisions. The Lemonade Lane simulation, a purpose-built five-round business simulation game, was played by 41 participants, randomly assigned to either a control group (n=23), who did not record their reasoning before seeing outcomes, or a research group (n=18), who documented their rationale before seeing each round's results. Three statistically significant differences were observed: (1) The logging group adjusted pricing more dynamically across rounds; (2) When prompted to explain their worst round post-game, loggers provided longer responses and named more distinct causal factors; (3) We used NLP analysis to compare each logger's notes before seeing the outcome with their explanations after the game for their worst round. We found that loggers showed a consistent tendency to shift from blaming outside factors to blaming themselves. Our findings suggest that integrating decision logging (CDL) into simulation-based games creates thinking patterns and evidence that can help facilitators mitigate the impact of hindsight bias during reflection. CDL leads to deeper, more detailed, and self-critical post-simulation reflection. This approach offers a practical step for designers seeking to enhance learning outcomes in game-based environments. These findings position CDL as a low-cost, scalable design intervention for practitioners seeking to amplify the pedagogical value of simulation-based games.

Keywords: Hindsight bias, Business simulation, Simulation-based learning, Debriefing, Decision logging

1. Introduction to Research

Business simulation games are widely used as learning tools. Learners gain value by reflecting on their decision-making. They understand why they made specific choices and what those choices reveal about their mental models. They can then use these reflections to adjust decision-making and mental models.

Most business simulations are structured in rounds. In each round, learners assess the simulation's current state, analyse available information, make decisions, and observe outcomes. Some outcomes are immediate; others, with a time delay, manifest several rounds later.

The reflective debrief, which typically follows completion of all rounds, represents the most pedagogically significant part of the simulation. Many facilitators use popular debriefing techniques, such as PEARLS Promoting Excellence and Reflective Learning in Simulation; Eppich and Cheng, (2015) and Schön's reflective practice framework (The Reflective Practitioner: How Professionals Think in Action).

During debriefing, instructors ask learners to articulate the rationale for their decisions. However, in most simulation games, the time between when learners make decisions and when they reflect on them ranges from several minutes to several hours. In simulations that require observable behavioural responses, assessors or facilitators can log observations. However, in simulation games, most reasoning takes place in the learner's mind and remains inaccessible.

Several cognitive factors reduce the accuracy of reflection after simulations. First, memory decay occurs as learners go through multiple rounds, causing earlier reasoning to fade or be overwritten. By the debriefing, participants often struggle to recall their original thought processes. Murre and Dros (2015) confirmed Ebbinghaus's findings on memory decay, supporting the forgetting curve.

The second major factor, hindsight bias, arises when knowledge of the outcome alters an individual's reasoning, leading explanations to appear more logical or consistent with the outcome than they did at the time of decision-making. Research by Fischhoff (1975) and Roese and Vohs (2012) shows that outcome knowledge does not just supplement memory; it actively reconstructs it.

Taken together, these effects call into question the reliability of post-hoc explanations during debriefing. If memory loss and outcome knowledge distort reported reasoning, we must re-examine the validity of simulation-based learning as an educational tool.

This paper proposes a simple intervention - Concurrent Decision Logging (CDL) to address this gap. At the end of each round of a simulation game, learners are prompted to record two things: (1) the rationale for the decision they just made, and (2) their confidence level in that decision. These real-time logs are captured before outcomes are known, creating a timestamped record of authentic reasoning.

During debriefing, learners are asked to revisit their decisions at the end of each round and compare their real-time logs with what they now recall and believe. This comparison anchors reflection in learners' actual thinking rather than reconstructed memory and provides the basis for a richer, more honest reflective conversation.

2. Objective

The objective of this paper is to evaluate whether Concurrent Decision Logging (CDL) reduces the distorting effects of hindsight bias on post-game reflection without undermining participant engagement with the simulation.

3. Our Research Questions

- Does concurrent logging reduce player engagement, as measured by logging volume and consistency across rounds?
- How does concurrent logging influence in-game decision-making?
- How does the reasoning documented in pre-outcome decision logs compare to the explanations players offer after seeing results?
- How does concurrent decision logging affect the quality and character of post-game reflection?

4. Experiment Construction

To study the effect of CDL on players' engagement, decision-making, and pre- and post-reasoning, a purpose-built simulation was designed based on the universally understood Lemonade Stall game, in which players run a lemonade stall for 5 business days.

Table 1: Game elements

Game Elements	Detail
Game	Lemonade Lane - a single-player browser-based simulation
Structure	5 rounds where each round represents one trading day
Decisions per round	Players can set price per glass Buy lemons, sugar, and ice Set a per-round marketing spend Create their lemonade recipe (lemons, sugar, ice per glass)
Key mechanic	Delayed impact of word-of-mouth (WOM): Recipe quality decisions in rounds 1–2 accumulate invisibly and surface as demand consequences from round 3 onward. Players cannot see the formula, so genuine uncertainty is preserved. Constant Rent: Players need to pay a constant rent of ₹200 even on days that they chose not to do any business.
Shock events	Round 3: a competitor opens nearby leading to price sensitivity in the market Round 4: heatwave causes demand to surge but supplier costs remain high Round 5: health trend leads early recipe quality to be rewarded or penalised
Financial goal	Grow ₹5,000 to at least ₹7,000 across 5 rounds
Framing	Player is given a loan and is accountable for returns. This is intended to increase decision engagement.
Deployment	Browser-based with no facilitator required. This ensures that the game is scalable to large samples.

Game Elements	Detail
Randomisation of players	Players assigned to research and control group through a randomiser. The playtesters were assigned code P01 to P42. They were then assigned to the research and control group using the Excel randomiser function =Randbetween(0,1).
Player selection	Player's representation: Gender: Female 20, Male 21

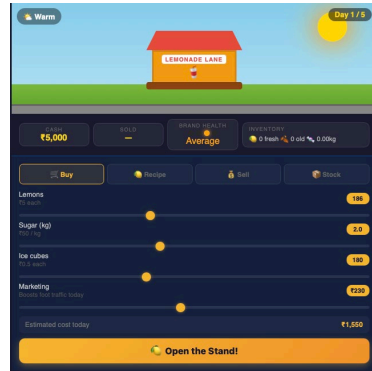


Figure 1: Recipe creation screen

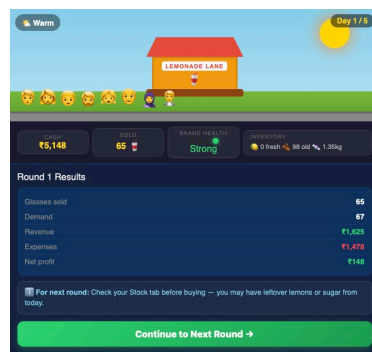


Figure 2: Round results

4.1 Dual-track Architecture

After each round:

Question 1 (research group only): What is driving your decisions today? What made you set this price, buy this quantity, and choose this recipe over the alternatives? (*Free text, 50 characters minimum*)

Question 2 (both groups): How much do you doubt that your decisions today will lead to positive results?

- 1: No doubt at all. I am confident in my plan today
- 2: Slight doubt. Mostly confident but not entirely certain
- 3: Mixed. Part confident, part uncertain about how today will go
- 4: More doubtful than confident. I am not sure this will work
- 5: Full of doubt. Something feels off, but I am going ahead anyway

End of game retrospective doubt rating:

Question 1 (both groups): Think back to Day X when you sold Y glasses against an expected Z. At the moment you clicked Open the Stand that morning and before you saw any results, how much did you doubt your decisions? (*Note: the X and Y values are automatically provided by the game and do not require recall from memory.*)

- 1: No doubt at all. I was completely confident in my decisions
- 2: Slight doubt. Mostly confident but with a small feeling of uncertainty

- 3: Mixed. Part confident, part uncertain
- 4: More doubtful than confident. I was not really sure my decisions were right
- 5: Full of doubt. Something felt wrong, but I went ahead anyway

End of game causal attribution (both groups):

Question 2 (both groups): In one sentence, what do you think caused it? (*Free text, 10 character minimum.*

Prompt context: "Focus on your worst round. What was the main reason results fell short of what you expected?")

4.2 Data Quality

42 participants played the game. Data from 41 participants were included in the analysis, with 18 in the research group (who documented their decision rationale before reviewing game outcomes) and 23 in the control group (who did not document their rationale before reviewing game outcomes). Only the first play session per participant was retained. One session was removed due to gibberish entries across all logging fields. Names and personally identifiable details were anonymised; players were assigned IDs P01–P41.

5. Results

5.1 Research Question 1: Does Concurrent Logging Reduce Player Engagement, as Measured by logging Volume and Consistency Across Rounds?

Observation 1: Log length remained stable across Rounds 1–4 (means: 22.6, 22.2, 21.9, 19.9 words) before a significant decline in Round 5 (16.1 words). The Round 5 compression is consistent with either end-of-game eagerness to see results or the reduced strategic complexity of a final round where inventory and brand position are largely fixed. Both explanations are compatible with the finding that all 18 players completed all five logs. Log quality may have compressed in Round 5 without engagement collapsing.

Table 2: Log length across rounds with standard deviations

	Round 1	Round 2	Round 3	Round 4	Round 5	Overall
Mean (words)	22.6	22.2	21.9	19.9	16.1	20.5
SD	15.4	16.5	18.7	18.8	11.7	—

Note: The high SDs reflect two outlier loggers (P14 and P38) who wrote substantially more than the rest of the group; without them, the means sit closer to 14–15 words per round throughout.

Interpretation: Concurrent Decision Logging (CDL) does not significantly lead to player fatigue or disengagement.

Observation 2: 100% of players in the research group completed the game.

Interpretation: Concurrent Decision Logging (CDL) did not reduce player motivation or engagement.

5.2 Research Question 2: How Does Concurrent Logging Influence in-game Decision-making Behaviour?

Observation 1: Players in the research group showed significantly larger intra-player variance in pricing decisions across rounds. This is the strongest between-group finding in the dataset. Both groups started at ₹25 (15 of 18 research, 14 of 23 control opened on Day 1 at exactly ₹25) and are virtually identical in mean price across rounds (research: ₹24.9; control: ₹24.8).

Table 3: Intra-player pricing variance by group

	Mean SD	Mean Variance
Research (n=18)	3.64	18.03
Control (n=23)	1.91	6.44

Intra-player variance across 5 rounds · Research players showed nearly 3× the pricing variance of control players
Mean price virtually identical (research ₹24.9 · control ₹24.8)

Observation 2: Response to round 3 - competitor opens nearby, leading to price sensitivity in the market

Table 4: Price response after round 3 where a competitor opens nearby, leading to price sensitivity in the market

Price response R2→R3	Research (n=18)	Control (n=23)
Reduced price ↓	38.9% (7 players)	26.1% (6 players)
Held price →	38.9% (7 players)	60.9% (14 players)
Raised price ↑	22.2% (4 players)	13.0% (3 players)

The key contrast: 60.9% of control players held their price unchanged in Round 3 despite the competitor shock. Only 38.9% of research players did. Research players were more likely to respond either by cutting prices to stay competitive (38.9%) or by actively pursuing an alternative strategy. Control players defaulted to holding.

Interpretation: Concurrent Decision Logging (CDL) leads players to move away from default anchors. Having been required to articulate their reasoning before each round, research players were more likely to revisit their prior decisions when conditions changed consciously. It can be inferred that logging made the reasoning explicit and, therefore, easier to challenge.

Logs for research players who named the competitor in their Round 3 log (6 of 18):

Table 5: Round 3 logs for research players who named the competitor

Player	Price change	Response type	Verbatim log extract
P01	₹25→ ₹25 (→)	Marketing response	"since a competitor is here, investing more on marketing"
*P02	₹15→ ₹18 (↑)	Undercut strategy	"cold day, competitor, i want to undercut the competition"
P06	₹25→ ₹25 (→)	Marketing response	"I am spending more on marketing because of the competitor"
P14	₹28→ ₹21 (↓)	Price cut	"cold weather and a nearby competitor will reduce customer demand and make people more price-sensitive"
P23	₹25→ ₹23 (↓)	Price cut	"Its cold and competitor is also there so reduced the price"
P25	₹25→ ₹20 (↓)	Price cut + marketing	"I'm lowering the price today as I have a competitor and cold weather"

* P02 entered Round 3 at ₹15 which is already below the default ₹25. Raising to ₹18 is consistent with an undercutting strategy relative to the competitor's assumed price.

5.3 Research Question 3: How Does the reasoning Documented in Pre-outcome Decision Logs Compare to the Explanations Players Offer After Seeing Results?

Observation: We found that players used hypothetical language in their post-game attributions (mean 0.83 operators per player; 8 of 18 players used at least one), while no hypothetical language appeared in any pre-outcome log. Words like "would have", "should have", "could have", and "maybe" were routinely used in the post-game attributions. This is a clear signal that knowledge of outcomes is changing players' reasoning.

Players who wrote operationally before the round ("going all out", "buying more lemons", "setting price at ₹25") shifted to counterfactual reasoning after ("maybe I should have changed the recipe", "I would probably take feedback on day 1").

Table 6: Comparison of pre-outcome logs vs post-game attribution

Player	Worst Round	Pre-outcome Log	Post-outcome Attribution	Interpretation
P07	D4	Log: "i am going all out. more spend on marketing higher price and great recipe."	Attribution: "i dont think i planned well hence i was underconfident"	Log shows confidence across three dimensions. Attribution discards all of them and replaces with one vague personal failing.
P14	D4	Log: "I increased stock because hot weather will bring higher customer demand... I set the selling price at ₹27 to cover higher supplier costs while still staying competitive despite the nearby competitor."	Attribution: "I underestimated demand during the heatwave and bought too little stock, which caused me to miss many potential sales."	The log shows deliberate stock decisions made in anticipation of high demand; the attribution frames the same event as a failure to anticipate demand.
P30	D5	Log: "warm weather - neutral opinion of customers- marketing can get more profit"	Attribution: "heat wave should have created more demand"	Log correctly identifies warm weather on D5. Attribution attributes failure to the heatwave which was in round D4 not D5. This shows that logging does not prevent factual distortion.
P38	D5	Log: "Not much. Slightly deflated on the demand lower. I think my decision on the ice hurt my profits. Keeping the purchases low today."	Attribution: "Maybe I didn't see the impact of the recipe on brand health. I don't think I had feedback on the recipe which could have helped. In the real world, I would probably take feedback on the recipe on day 1 itself, and adjust accordingly."	Attribution introduces a causal mechanism not mentioned in any log. The player links poor results to a feedback gap and transfers the learning to real-world practice. This attribution is the clearest example of CDL producing deep reflection.

Interpretation: The shift from operational to counterfactual and hypothetical language is the clearest signature of hindsight bias. Players could not have written "should have" before the outcome, as they did not yet know it. Once they knew it, they constructed an alternative past.

Important: This also shows that logging alone does not guarantee protection against hindsight bias. However, the log creates a trail that the facilitator can use during post-game reflection to counter hindsight bias.

For example: A facilitator who shows P07 their own log alongside their attribution can ask: 'You said you were going all out across three dimensions: marketing, price, and recipe. You then said you didn't plan well. What changed between those two statements?' That question is only possible because the log exists.

5.4 Research Question 4: How Does Concurrent Decision Logging Affect the Quality and Character of Post-game Reflection?

Observation: The research group wrote roughly 53% more text when asked to reflect on their worst round. The control group provided brief, transactional answers, whereas the research group provided narrative explanations. The research group averaged 20 words, compared with 13.1 words for the control group.

The most significant between-group difference in the dataset is causal count: research players named significantly more distinct causes in their attributions, with 77.8% naming multiple causes versus 39.1% in the control group.

Table 7: Frequency analysis of causal vocabulary in final attributions

Category	Word / Phrase	Research count	Research players	Control count	Control players
Causal connectors	because	1	1 of 18	0	0 of 23
	hence	1	1 of 18	0	0 of 23
	which caused	1	1 of 18	0	0 of 23
	that caused	1	1 of 18	0	0 of 23
	that's why	1	1 of 18	0	0 of 23
	caused by	0	0 of 18	1	1 of 23
	since	1	1 of 18	0	0 of 23
Self-evaluative	should have	3	3 of 18	1	1 of 23
	would have	2	2 of 18	0	0 of 23
	overconfident	1	1 of 18	0	0 of 23
	underestimated	1	1 of 18	0	0 of 23
	ignored	1	1 of 18	0	0 of 23
	mistake	1	1 of 18	0	0 of 23
	could have	1	1 of 18	1	1 of 23
	didn't	1	1 of 18	1	1 of 23
	lack of	1	1 of 18	2	2 of 23
	did not	0	0 of 18	2	2 of 23
Decision-specific	recipe	9	4 of 18	1	1 of 23
	stock	3	3 of 18	1	1 of 23
	inventory	2	1 of 18	0	0 of 23
	margins	1	1 of 18	0	0 of 23
	waste	1	1 of 18	0	0 of 23
	lemons	1	1 of 18	0	0 of 23
	marketing	2	2 of 18	2	2 of 23
	price	2	2 of 18	2	2 of 23
	pricing	0	0 of 18	3	3 of 23
	brand	1	1 of 18	3	2 of 23
	planning	1	1 of 18	1	1 of 23
External / contextual	demand	5	4 of 18	4	4 of 23
	weather	3	3 of 18	2	2 of 23

Category	Word / Phrase	Research count	Research players	Control count	Control players
	rainy	2	2 of 18	1	1 of 23
	competition	2	2 of 18	1	1 of 23
	customers	2	1 of 18	1	1 of 23
	market	2	4 of 18	2	4 of 23
	heatwave	1	1 of 18	0	0 of 23
	information	0	0 of 18	1	1 of 23
	setup	0	0 of 18	1	1 of 23

Three patterns stand out across the full table.

- Research players hold a clear advantage on causal connectors: “because”, “hence”, “which caused”, “that’s why” all appear exclusively in research attributions, meaning only research players explicitly linked a decision to an outcome rather than simply naming both.
- On self-evaluative vocabulary, “overconfident”, “underestimated”, “ignored”, and “mistake” are research exclusive.
- On decision-specific terms, “recipe” is the starkest difference with 9 occurrences across 4 research players versus 1 in the entire control group. Control players show higher use of “pricing” and “brand”, but these tend to appear as vague category labels rather than diagnosed mechanisms.

Interpretation: Concurrent logging is associated with richer, more causally specific post-game reflection. Research players not only wrote more but also identified more distinct causes and used more precise diagnostic language when explaining their worst round.

Overall observation: The game outcomes showed no significant difference between the research and control groups.

Table 8: Game outcomes by group

Outcome	Research (n=18)	Control (n=23)
Final cash balance	₹4,300	₹4,407
Final game score	49.7	52.7

Interpretation: Concurrent Decision Logging (CDL) did not change game outcomes, so logging itself did not improve players’ performance.

6. In Conclusion

Within the constraints of this study, four practical conclusions can be drawn for designers considering using Concurrent Decision Logging:

- Logging did not reduce completion rates or log volume, so designers need not treat participant fatigue as a reason to exclude it from in-game reflection.
- Logging made no measurable difference to game outcomes (final cash, final score) in this study; it should not be adopted on the expectation that it will improve in-game performance.
- Logging players produced markedly richer, more causally specific reflections on their outcomes. This is a pattern consistent with, though not direct evidence of, an improved potential to transfer learning beyond the game.
- Logging supplies facilitators with a documented, timestamped account of pre-outcome reasoning that can be directly compared with post-game attribution. The log acts as a practical, low-cost mechanism for making hindsight bias visible and contestable during debrief.

7. Future Research

- The study must be extended to different simulation games (including face-to-face and analogue formats, in which structured written reflection would replace digital logging) to determine whether the results can be replicated.
- Testing whether showing players their log during debrief actually improves learning outcomes, which the current study does not measure.
- Whether the depth and specificity of pre-outcome logs predict the quality of post-game attribution, and if so, what prompts would maximise that effect.
- How can facilitators use pre-outcome logs and post-game attributions to surface and challenge distorted retrospective accounts, and does doing so measurably improve the quality of post-simulation learning?

Our ultimate aim would be to provide a research-informed framework of debriefing and reflection that can make simulation games effective learning instruments beyond their engagement and experiential value.

The majority of work in the field of game-based learning and simulations focuses on the more exciting aspects of mechanics, dynamics, and aesthetics. We propose a shift toward debriefing-focused game-based learning and simulation design, demonstrating that without structured strategies for reflection and the integration of learning, games and simulations cannot fully realise their pedagogical potential.

8. Acknowledgements

The authors thank all participants who gave their time to play Lemonade Lane. Special thanks to Numair Antule, who managed the web server and the technical infrastructure required to deploy the game and collect player data.

Ethics Declaration: No ethical clearance was required for this research.

AI Declaration: The following tools were used in the writing of this paper:

- Grammarly for editing and text proofing
- Claude.ai for the NLP analysis
- Claude.ai for statistical analysis

References

- Cheryl I. Johnson, Richard E. Mayer, Applying the self-explanation principle to multimedia learning in a computer-based game-like environment, *Computers in Human Behavior*, Volume 26, Issue 6, 2010, Pages 1246-1252, ISSN 0747-5632.
- Eppich, W. and Cheng, A. (2015) "Promoting Excellence and Reflective Learning in Simulation (PEARLS): Development and Rationale for a Blended Approach to Health Care Simulation Debriefing", *Simulation in Healthcare*, Vol 10, No. 2, pp 106–115.
- Fischhoff, B. (1975) "Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty", *Journal of Experimental Psychology: Human Perception and Performance*, Vol 1, No. 3, pp 288–299.
- Harold F. O'Neil, Gregory K.W.K. Chung, Deirdre Kerr, Terry P. Vendlinski, Rebecca E. Buschang, Richard E. Mayer, Adding self-explanation prompts to an educational computer game, *Computers in Human Behavior*, Volume 30, 2014, Pages 23-28, ISSN 0747-5632.
- Motavalli A, Nestel D. Complexity in simulation-based education: exploring the role of hindsight bias. *Adv Simul (Lond)*. 2016 Jan 11;1:3. doi: 10.1186/s41077-015-0005-7. PMID: 29449972; PMCID: PMC5796605.
- Murre, J.M.J. and Dros, J. (2015) "Replication and Analysis of Ebbinghaus' Forgetting Curve", *PLOS ONE*, Vol 10, No. 7, e0120644.
- Roese, N.J. and Vohs, K.D. (2012) "Hindsight Bias", *Perspectives on Psychological Science*, Vol 7, No. 5, pp 411–426.
- Schön, D.A. (1983) *The Reflective Practitioner: How Professionals Think in Action*, Basic Books, New York.
- Tobias Alf, Marieke de Wijse, and Friedrich Trautwein, The Role of Reflection in Learning with Simulation Games – A Multi-Method Quasi Experimental Research, *Volume 54, Issue 6*.
- Yang, X., Rahimi, S. Prompting self-explanation in different gameplay contexts: effects on response quality and content learning. *Education Tech Research Dev* (2026).