

Simulation and conceptual change: an approach based on the analysis of student discourse

Santiago Porras Alfonso¹ , Julio César Puche Regaliza¹ , Athénaïs Sauvée² , Silvia Casado Yusta¹ 

¹Department of Applied Economics, University of Burgos, Spain, ²Department of Private Law, University of Burgos, Spain.

How to cite: Porras Alfonso, S.; Puche Regaliza, J. C.; Sauvée, A.; Casado Yusta, S. (2026). Simulation and conceptual change: an approach based on the analysis of student discourse. In: 12th International Conference on Higher Education Advances (HEAd'26). Valencia, 15-18 June 2026. <https://doi.org/10.4995/HEAd26.2026.21681>

Abstract

Analysing how students reason when faced with probabilistic problems is key to understanding and improving teaching and learning processes in contexts of uncertainty. Although simulation has proven effective in increasing correct answers, there is less evidence on whether these changes transform the underlying reasoning. This paper analyses the conceptual change at the behavioural, probabilistic and reasoning levels of students after a simulation process using the Monty Hall problem. Using statistical techniques, natural language processing and semantic network analysis, it is concluded that there is a conceptual change at all three levels. Pre-simulation reasoning is dominated by simple, equiprobable heuristics, while post-simulation reasoning is dominated by normative heuristics, demonstrating that the simulation process has a positive effect on the internalisation of probabilistic concepts, as well as a reorganisation of explanatory discourse.

Keywords: Monty Hall dilemma; Natural Language Processing; Semantic network analysis; Cognitive heuristics, Probabilistic reasoning;

1. Introduction

One of the main tasks of teachers is to mark exercises, practical work or writing exams. The answers given by students are sometimes brilliant, sometimes incorrect, leading teachers to ask themselves: What was this student thinking to reach that conclusion? Sometimes the assessment of learning focuses solely on the final result: whether the answer is right or wrong. Although this method is useful, it does not allow us to understand how reasoning is structured or what ideas guide the decision-making process.

Strategies like Monte Carlo simulation, a computational process using repeated random trials, enhance understanding of these problems. The academic literature indicates that these tools generate behavioural changes, such as a higher proportion of correct answers. However, it is less common to examine whether these changes transform the way students justify their decisions. This raises a key question: does a change in response necessarily imply a change in reasoning? This study aims to analyse students' conceptual change from a multilevel perspective, combining behavioural, probabilistic and reasoning measures. To evaluate these three levels, the following hypotheses are proposed:

H1: Do significant changes occur in students' behavioural decisions after the simulation?

H2: Are there changes in students' probabilistic responses after the simulation?

H3: Are there changes in the students' reasoning structures and the heuristics used after the simulation?

The Monty Hall problem is used as a controlled test case due to its well-documented cognitive difficulty and its ability to elicit systematic reasoning biases. Its simplicity allows isolating changes in reasoning without introducing additional contextual complexity and it allows the three proposed levels of analysis to be clearly observed. Statistical techniques, natural language processing and semantic network analysis will be used to test the hypotheses.

2. Related Work

The Monty Hall problem is a classic case of probabilistic reasoning under uncertainty introduced by Selvin (1975). It poses a dilemma of choosing between three doors, two of which contain a goat as a prize and the third a luxury car. The contestant chooses a door and then the presenter reveals one of the unchosen doors, which contains a goat. The contestant is then faced with the dilemma of whether or not to change doors. The logical solution to the problem is to switch doors, as this increases the probability of winning the prize to two-thirds. However, a large number of contestants continue to use flawed reasoning, even after the reasoning behind the problem has been explained, as demonstrated in the work of Krauss & Wang (2003).

The answers given by people who are presented with the problem correspond to systematic patterns of reasoning known as cognitive heuristics. The most relevant heuristics in the Monty Hall problem that have been scientifically documented are: The equiprobability heuristic assumes that, after opening one door, the probability is equal for the two remaining, with terms such as 'half' or 'fifty-fifty'. The initial fixation heuristic maintains the first choice based on intuition or resistance to change, without considering probability using terms such as "first" or "stick". The heuristic of ignoring the presenter's condition does not consider that the open door is not random, but conditioned by the location of the car. The normative heuristic analyses the problem probabilistically, using terms such as "probability", "one third", "two thirds", and is

uncommon before explaining the reasoning. The hesitant heuristic, which expresses indecision with terms such as "it doesn't matter" and the simulation heuristic, where the answer is justified by experience, using terms such as "simulate".

Simulation is a tool that improves comprehension in probabilistic reasoning problems such as Monty Hall. This is because it allows the problem to be worked on repeatedly and the regularities in the results to be observed (Puche Regaliza et al., 2025). These studies indicate that the simulation process increases behavioural change in terms of changing doors and also in terms of probabilistic response.

In the Monty Hall dilemma, performing a simulation process helps to move from simple intuitive heuristics to more complex ones, but this does not imply reaching normative reasoning. It is therefore essential to analyse the reasoning process expressed in the participants' language to determine whether a change in reasoning occurs after the simulation.

3. Methodology

3.1 Study context and experimental design.

The experiment carried out follows three phases: In the first (PRE), students are presented with the Monty Hall dilemma and asked to answer whether or not they would change doors, the probability they believe there is of success, and to briefly explain their reasoning. Next, in the second simulation phase, they are given 10 minutes to simulate Monty Hall with a tool designed by the authors (Puche Regaliza et al., 2025). This tool allows them to play individual games or simulate a certain number of them. It also includes three player profiles to simulate: one who always changes doors, one who never changes, and one who changes every other time. Finally, in the third phase (POST), they are asked to answer the same questions again.

The sample contains the responses of 85 first-year students studying for degrees in business administration and management and finance and accounting. The aim of this simulation is to highlight the importance of information in decision-making. The samples are not paired for data protection reasons, which is why the study is conducted at group level.

3.2 Analysis techniques and metrics applied.

The following techniques are used to evaluate the hypotheses proposed. First, to see whether or not there is a behavioural change, the Chi-square statistical technique (Agresti, 2012) is used to determine whether there are statistically significant differences between those who choose to change doors before the simulation and those who choose to change after it. For the second hypothesis, to check whether or not there are differences in the probabilistic responses, the Mann Whitney-U technique (Mann & Whitney, 1947) is used to see if there are statistically

significant differences between the probabilities indicated by the students before and after the simulation. This test is suitable for unpaired samples.

The third hypothesis, to assess whether or not there is a change in the students' reasoning after the simulation and a change in heuristics, requires natural language processing (NLP). This process is carried out for both the responses before and after the simulation in order to observe whether there is a change in reasoning. In a first step, the textual responses were preprocessed (lowercasing, removal of punctuation and stopwords), lemmatised, and filtered by grammatical category using UDPipe (Straka et al., 2016). As the responses were short texts, the n-gram technique (Manning & Schutze, 1999) was selected to perform the lexical analysis (Gries, 2016), which allows us to see the linguistic patterns that emerge in the responses and subsequently relate them to cognitive heuristics. The next step in the analysis is to construct the pre- and post-co-occurrence networks (Evert, 2004), where the nodes represent the linguistic terms and the edges represent the relationships between them. Once constructed, the communities are calculated using the Louvain technique (Blondel et al., 2008) and semantic networks are constructed (Steyvers & Tenenbaum, 2005), assigning the nodes of the network to the cognitive heuristic to which they belong, based on adapted lexical dictionaries (Vosniadou, 1994) as described in section 2. This process allows us identify the responses with the cognitive heuristics. The fact that a term belongs to a heuristic dictionary does not mean that it cannot be associated with another heuristic. This is because the dictionary assignment reflects semantic meaning, while final heuristic roles are determined by each term's position and connections within the network. As a final step, the network metrics are calculated: degree, centrality and betweenness to characterise the structure of reasoning and the distribution of heuristics.

4. Results

This section shows the results obtained. All data analyses were performed in RStudio. The results of the Chi-square tests for H1 yield a Chi-square value of 17.942 with a p-value of 2.99×10^{-5} which is smaller than 0.05, indicating that there are statistically significant differences in the action of changing doors or not before and after the simulation.

The Mann-Whitney-U test for H2 indicates a p-value of 0.022, which is less than 0.05. Therefore, there are statistically significant differences in the probabilities indicated by the students before and after the simulation.

Figure 1 shows variations before and after simulation. The percentage of students who would change doors increased from 38% to 72%. For probabilistic reasoning, students choosing the correct probability of 0.67 increased from 32.1% to 49.4%.

Figure 2 shows the resulting semantic networks for evaluating H3. The nodes represent the concepts used by the students, the edges represent the co-occurrences that establish that the terms appear in the same answer, and their thickness indicates the number of times they appear related. The size of the nodes indicates the degree, which is the number of times the term appears. The shape of the node indicates the community to which it belongs, and the colour is associated with the cognitive heuristic assigned by the dictionaries.



Figure 1 Proportion of students who change doors before and after the simulation, and estimated probability of success before and after.

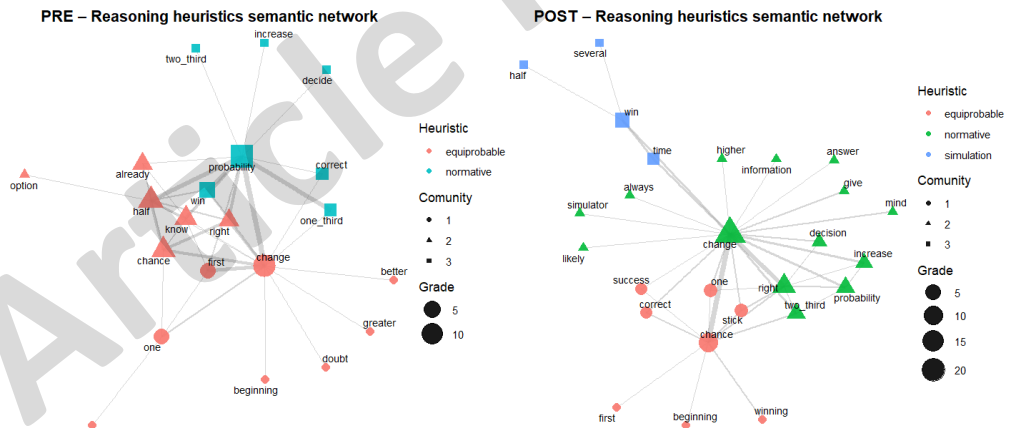


Figure 2 Comparison of semantic networks before the simulation (PRE) and after (POST)

In the semantic network pre-simulation, normative and equiprobable heuristics appear. Equiprobability dominates with 14 nodes, an average degree of 4.07 and betweenness of 0.058,

as seen in Table 1. This shows students favor equiprobable reasoning. The "change" node has the highest degree and betweenness (14 and 0.498), as shown in Table 2. Its position and connections make it a hub node that interconnects terms rather than unifying reasoning.

Table 1 Network metrics.

Network	Heuristic	Nodes	Mean degree	Mean betweenness
PRE	Equiprobable	14	4.07	0.058
	Normative	7	3.29	0.051
POST	Equiprobable	8	2.88	0.031
	Normative	14	3.50	0.060
	Simulation	4	2.00	0.040

Table 2 Top 5 nodes in each network based on their degree.

PRE			POST		
Node	Degree	Betweenness	Node	Degree	Betweenness
Change	14	0.498	Change	20	0.792
Probability	12	0.354	Chance	10	0.248
Half	8	0.118	Right	8	0.037
Chance	7	0.050	Probability	4	0.001
Know	6	0.018	Two third	4	0.004

The normative heuristic appears in blue, with 7 nodes, an average degree of 0.329 and a betweenness of 0.051 indicating that certain students demonstrate some intuition towards the problem before the simulation. The central node of this network is "probability," which appears associated with normative terms such as "two thirds" or "increase," which indicates that a small group of students has understood the problem since they are in a peripheral position. However, its strongest relationships are established with equiprobable heuristic terms such as "half," indicating that, it cannot be established that there is strong normative reasoning. Three communities and two heuristics appear in this network; they are independent entities. A community expresses a set of nodes that are strongly related to each other, and heuristics are the cognitive reasoning associated with those terms and the terms with which they are related.

In post-simulation network, a structural change can be observed. Normative heuristic dominates the network in a central position with 14 nodes, an average degree of 3.50, and a betweenness of 0.06. The central node is 'change', with a degree of 20 and a betweenness of 0.792, which changes community and heuristics, now grouping normative terms such as 'two-thirds', 'information' and 'increase'. This means that it is now the center of the students' reasoning and connects the different communities and heuristics. In contrast, the node "probability," which was central in pre-simulation, now appears displaced which indicates the abandonment of a more generic term in favour of more concrete normative terms. The equiprobable heuristic loses its central place, reducing its number of nodes to 8 and its betweenness to 2.88, indicating that

students abandon the heuristic in favour of the normative, although this shows that there are students who still do not understand the problem after the simulation. Finally, a simulation heuristic appears with a small reference with 4 nodes, but high betweenness, indicating that students include the simulation process in their reasoning. The simulation heuristic is linked to the normative and is unrelated to the terms of equiprobability, which reinforces the fact that simulation increases normative reasoning. All in all, it can be concluded that after the simulation there is a transition from equiprobable to normative reasoning, accepting H3.

5. Conclusions and Limitations

This paper analyses students' conceptual change after a simulation process at three levels: behavioural, probabilistic and reasoning. While changes at the first two levels are well documented and confirmed by this study, changes in reasoning have received less attention. The results show that, after the simulation process, the proportion of students who use normative reasoning based on the simulation increases and are able to give the correct answer. Even so, there are students who do not manage to reach this level of reasoning and remain at the level of equiprobable cognitive heuristics. Methodologically, it is demonstrated that NLP analysis with networks is valid for detecting changes in the cognitive process of students after simulation.

Pedagogically, the results suggest simulation-based activities effectively promote deeper reasoning in probabilistic contexts. Instead of focusing solely on correct answers, instructors can use simulation to encourage students to reflect on patterns, compare strategies, and articulate reasoning. Also, analysing students' explanations through simple textual tools may help educators identify misconceptions and adapt instruction.

One limitation is the sample consists of students with similar profiles. Another is the samples are not paired, preventing more complex analyses. Finally, student language differs from proposed dictionaries, necessitating adaptation. Future research should expand the sample to students from various degrees and courses. It is desirable to collect data from paired samples to identify student profiles and heuristics. Including different problems is also proposed to analyze student responses in various scenarios.

Acknowledgements

This work was supported by research grants from the University of Burgos and IFIE in the project "Agent-based simulation tool for gamification treatment: Analysis of the learning process of complex concepts through artificial intelligence" Call for Grants to Recognised Teaching Innovation Groups for developing teaching materials for 2025 and 2026.

References

- Agresti, A. (2012). *Analysis of Ordinal Categorical Data: Second Edition*. Analysis of Ordinal Categorical Data: Second Edition, 1–396. <https://doi.org/10.1002/9780470594001>
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Evert, S. (2004). *The Statistics of Word Cooccurrences Word Pairs and Collocations*.
- Gries, S. T. (2016). Quantitative corpus linguistics with R: A practical introduction. *Quantitative Corpus Linguistics with R: A Practical Introduction*, 1–274. <https://doi.org/10.4324/9781315746210>
- Krauss, S., & Wang, X. T. (2003). The Psychology of the Monty Hall Problem: Discovering Psychological Mechanisms for Solving a Tenacious Brain Teaser. *Journal of Experimental Psychology: General*, 132(1), 3–22. <https://doi.org/10.1037/0096-3445.132.1.3>
- Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Source: The Annals of Mathematical Statistics*, 18(1), 50–60.
- Manning, C., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. <https://philpapers.org/rec/MANFOS>
- Puche Regaliza, J. C., Porras Alfonso, S., Casado Yusta, S., & Pacheco Bonrosto, J. (2025). Impact of Gamification to Understanding Bayes' Theorem: A Study Using the Monty Hall Dilemma Simulation. *Dirección y Organización*, 86(86), 16–28. <https://doi.org/10.37610/86.696>
- Selvin, S. (1975). On the Monty Hall Problem. *The American Statistician*, 29(3), 134–134. <https://doi.org/10.1080/00031305.1975.10479121>
- Steyvers, M., & Tenenbaum, J. B. (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, 29(1), 41–78. https://doi.org/10.1207/S15516709COG2901_3
- Straka, M., Hajič, J., & Straková, J. (2016). UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 4290–4297. <http://hdl.handle.net/11234/1-1548>
- Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, 4(1), 45–69. [https://doi.org/10.1016/0959-4752\(94\)90018-3](https://doi.org/10.1016/0959-4752(94)90018-3)