

## BELLMAN-OPTIMAL DECISIONS AND EXPERT INTUITION

Iliia Atanasov, George Mengov<sup>✉</sup>, Anton Gerunov

*Received on February 13, 2025*

*Presented by K. Atanassov, Member of BAS, on March 25, 2025*

### Abstract

Engineers and other professionals believe in rationality and seek to make optimal choices most of the time. Their expert intuition is developed over years of education and practice, enhanced by various decision-support tools. Yet being human, they are prone to emotional biases leading away from the best judgement and action. Here we report statistically significant deviations from the Bellman optimality principle by participants in a lab experiment about managing an abstract production system. We find that when the supply of a key resource diminishes and people are surrounded by others in the same position, they perform below a weak form of the Bellman-optimal criterion. In contrast, one's choices become much more successful when additional expert information is made available.

**Key words:** Bellman optimality principle, decision making, expert intuition

**Introduction.** While Artificial Intelligence is here to stay, it still has essential cognitive deficiencies making its use problematic [1, 2]. Human judgement remains indispensable in high-stakes circumstances, such as nuclear power plant operation, air traffic control, chemical and metallurgical facilities operation, medical treatment, decisions on war and peace, policy design and implementation, and

---

This study is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project SUMMIT BG-RRP-2.004-0008-C01.

<https://doi.org/10.7546/CRABS.2025.05.07>

many more. Decision making under risk and uncertainty in any domain is the privilege and responsibility of humans, preferably of seasoned professionals. It is believed that the highest form of mature reasoning is the expert intuition [3, 4], which is more advanced than logical reasoning and is developed over many years of education and experience accumulation. At the same time, people in their actions are influenced by well-studied cognitive biases and emotional influences that render their decisions suboptimal.

In this study we report findings from a lab experiment, in which participants had to make repeated choices aimed at utility maximization. They were asked to deal with a resource abstract enough to represent a multitude of possibilities in many industries and branches of the economy and society. The experimental design allowed for a comparison of the actual human behaviour with two different implementations of the Bellman optimality principle. We sought to identify some of the factors leading to inferior decision making by people, and the extent to which it deviated from being Bellman-optimal, which served as a ‘golden standard’.

**Materials and methods. *Experimental setup.*** The experiment was conducted in a computer laboratory where groups of 14–18 participants took part in a game, aimed at maximizing the abstract commodity ‘omnium bonum’. Translated from Latin, that means simply ‘a good for everyone’ and the peculiar name was needed to induce a non-emotional and cognitively neutral response by all people. The game was computer-based and consisted of 20 rounds with equal content but different parameters. Each round began by showing offers of varying quantity of omnium bonum by four suppliers. It was known that they might not act as promised and might deliver a higher or lower amount. A participant could choose only one offer. After a delivery, the player compared it with what had been offered initially and expressed their level of satisfaction or disappointment on a scale between +4 and –4. Their choice of supplier and the self-assessed emotion were automatically sent to every other participant’s computer and put in a queue. Those messages were displayed in a screen corner one by one every two seconds, highlighted in a red frame for 0.4 seconds, to attract attention. The point was to inform the player what everybody else was doing, and how satisfied they were with each supplier. The game lasted about 20 minutes and ended by paying each participant a small sum of money in proportion to their accumulated omnium bonum.

To test human propensity for sub-optimal decisions, we designed four experimental conditions. In two of them the quantities of omnium bonum rose steadily to the end, thus rewarding experience buildup. In the other two, the last five rounds offered diminished quantities, simulating a production decrease. We hypothesized that in the former two conditions people would act closer to Bellman optimality due to their own experience, supported by the collective experience of all other participants. In contrast, the latter two conditions could only stir dissatisfaction and anger, amplified by the lab virtual social network. Naturally,

less optimal decisions could be expected there.

Dividing the experimental conditions orthogonally, into half of them we introduced additional relevant information. That consisted of data about the total production of omnium bonum in the last round, and an unbiased forecast about the current round. The purpose was to provoke the players' expert intuition and lead them to closer-to-optimal decisions. A reasonable hypothesis was that people in the condition with monotonic growth and additional relevant data would achieve the best results.

All elements of the experimental design were tailored to the sample of participants. Those were students of Economics, Business Administration, and Public Administration at Sofia University "St. Kliment Ohridski". They had taken or were taking courses in calculus, algebra, probability and statistics, microeconomics, and macroeconomics. All of that equipped them with sufficient knowledge to claim the level of expert intuition required by the experiment. A total of 131 people played the game and were distributed approximately evenly over the four experimental conditions. Further technical details are given in [5].

**Theory and implementation.** Here we provide the necessary theoretical framework for understanding how the Bellman optimality principle was used in the experiment. First, we describe the general case. Let us define a directed multigraph as an ordered triple  $G = \langle v, E, \phi \rangle$ , where  $v = \{v_1, v_2, \dots, v_n\}$  denotes the finite set of all  $n$  nodes,  $E$  denotes the finite set of all edges and  $\phi : E \rightarrow v \times v$  denotes the edge assignment function that assigns each edge to an ordered pair of nodes.

Here, set  $E = \{e_{i,i+1}^k \mid \text{edge } k \text{ connects node } i \text{ with node } i + 1\}$ . The edge assignment function assigns the edges to the nodes as  $\phi(e_{i,i+1}^k) = (v_i, v_{i+1})$  for  $\forall i \in \{i = 1, 2, \dots, n - 1\}$  and  $k = \{1, 2, \dots, m\}$ , where  $m$  denotes a common number of edges between every node  $v_i$  and  $v_{i+1}$ . The structure of this graph is such that  $m$  edges will connect the node  $v_i$  with the node  $v_{i+1}$ , but there are no direct connections between node  $v_i$  and node  $v_{i+j}$  for  $j \neq 1$ . If a reward is assigned to each edge by a reward assignment function  $r_{i,i+1}(e_{i,i+1}^k)$ , one can traverse the multigraph from each node  $v_i$  to the final node  $v_n$  in an optimal way by following Bellman's principle of optimality:

*An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.*

Then the optimal reward that one can achieve starting from node  $v_i$  and reaching  $v_n$  must satisfy the following Bellman equation:

$$V_i(v_i, v_{i+1}) = \max \{r_{i,i+1}(e_{i,i+1}^1), \dots, r_{i,i+1}(e_{i,i+1}^m)\} + \hat{V}_{i+1}(v_{i+1}, v_{i+2})$$

$$\forall i \in \{i = 1, 2, \dots, n - 1\},$$

where  $V_i(v_i, v_{i+1})$  denotes the value function of the choice between which edge to traverse when moving from node  $i$  to node  $i + 1$  and  $\hat{V}_{i+1}(v_{i+1}, v_{i+2})$  denotes the value function of the next decision. Because of the specific structure of the graph, the available future rewards are independent of the current choice. Then the optimal reward from all choices that one can achieve when traversing the graph from node  $v_i$  to node  $v_n$  can be calculated as the sum of all rewards from the independent optimal choices between the initial and the final node.

Our experiment can be viewed as a special case of the multigraph described above. We introduce the following notation. By  $t = 1, \dots, 20$  we denote the experimental rounds. The suppliers are  $S = \{A, B, C, D\}$ . In round  $t$ , supplier  $i$  is promising a quantity of omnium bonum  $B_{i,t}^o$ . The actual delivery is  $B_{i,t}^d$ . To align with the Bellman optimality context, we denote as  $V_t(B_{i,t}^d)$  the value function of the choice in  $t$ . Each agent is characterized by a linear utility function  $u_t$ .

The experiment can be viewed as a directed multigraph with  $T = 20$  nodes plus a final node for result calculation ( $n = 21$ ). Every node of the first twenty represents a round  $t$  and contains information about the omnium bonum gathered already. There are four directed edges between every two neighbouring nodes  $t$  and  $t + 1$ . Each edge represents the delivered quantity by the  $i$ -th supplier in  $t$ . Figure 1 illustrates the transition process.

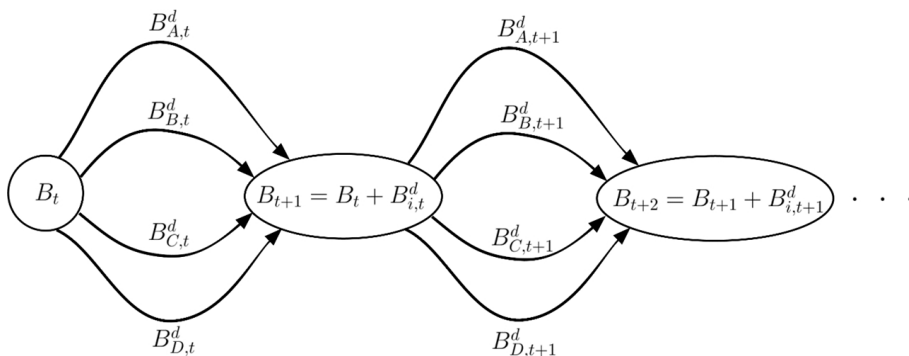


Fig. 1. Transition between rounds represented as a directed multigraph

In this setting, the player seeks to solve the following optimization problem:

$$(1) \quad \max \sum_{t=1}^{20} u_t(B_{i,t}^d) .$$

If all quantities, potentially delivered by each supplier in each round are known, the amount that maximizes Eq. (1), satisfies the following Bellman equation:

$$(2) \quad V_t(B_{i,t}^d) = \max\{B_{A,t}^d, B_{B,t}^d, B_{C,t}^d, B_{D,t}^d\} + \hat{V}_{t+1}(B_{i,t+1}^d) \quad \forall t.$$

Here,  $\hat{V}_{t+1}(B_{i,t+1}^d)$  is the maximum value gained from the best choice in the next round  $t + 1$ . Because a future state does not depend on the previous states, Eq. (2) can be solved for each round recursively, starting at 20 and going back to 1. A policy function can then be used for each round to map the maximum value to the corresponding choice. This theoretically best solution, however, is achievable only if one has full knowledge of the experimental design. Our participants were not in that position, yet this solution has practical value for various other contexts. For example, in semiconductor chip production or any other high-tech robotized production the operator must have at their disposal full amount of the relevant information for every stage of the process.

In our experiment though, that remained an unattainable golden standard. A less demanding and more realistic Bellman-optimal outcome can be a probabilistic one. We denote as  $p_{i,+}$  the average probability for each supplier  $i$  to deliver more than what was promised across all rounds in the experimental condition. Similarly,  $p_{i,-}$  is the probability that less than what was promised will be delivered. Finally,  $p_{i,o}$  is the probability to deliver exactly as promised. In addition, the agent knows the average deviation  $\delta_i$  from the promised quantity for each supplier  $i$ . While the latter circumstance may not be entirely realistic at the early stages of the game, with experience and additional information from the other participants in the network, one has the opportunity to develop such intuition as the game progresses. It is known that in such circumstances people are Bayesian-optimal intuitive statisticians [6]. Now, the agent's utility function is:

$$u(B_{i,t}^o) = p_{i,o}B_{i,t}^o + p_{i,+}(1 + \delta_i)B_{i,t}^o + p_{i,-}(1 - \delta_i)B_{i,t}^o.$$

Thus, the following optimization problem can be formulated:

$$(3) \quad \max \sum_{t=1}^{20} u_t(B_{i,t}^o).$$

Assuming perfect knowledge of the above probabilities, the optimal sequence of choices satisfies the following Bellman equation:

$$(4) \quad V_t(B_{i,t}^o) = \max \{u_t(B_{A,t}^o), u_t(B_{B,t}^o), u_t(B_{C,t}^o), u_t(B_{D,t}^o)\} + \hat{V}_{t+1}(B_{i,t+1}^o) \forall t.$$

Eq. (4) can be solved as Eq. (2), with the addition that the probabilistic account uses also the average deviations  $\delta_i$ .

**Results.** To assess how close to optimal people's decisions were, we conducted statistical comparisons. Conceptually, such human-generated data are assumed to be Gaussian-distributed and therefore the Student Test is an appropriate tool to use. However, as the samples were composed of 20 elements each, which might be too few, it is a good practice to use also a non-parametric test for the medians. The most suitable here is a version of the Wilcoxon Rank Sum

Test, which takes into account the order of observations. Thus, we computed the averages and medians for each round over all participants and compared them with the two Bellman-optimal models. As Table 1 shows, both tests produced identical results with respect to statistically significant differences.

T a b l e 1

Comparison of the median human behaviour with two Bellman-optimal solutions

Experimental Condition	Wilcoxon Rank Sum Test W-statistic ( <i>p</i> -value)		Student Test T-statistic ( <i>p</i> -value)	
	Full Information Model	Probabilistic Information Model	Full Information Model	Probabilistic Information Model
Growth	<b>-3.924</b> <b>(0.0000)</b>	-1.394 (0.1632)	<b>-6.275</b> <b>(0.0000)</b>	-1.5311 (0.1422)
Growth, Available Expert Data	<b>-3.8277</b> <b>(0.0001)</b>	0.1812 (0.8562)	<b>-7.1053</b> <b>(0.0000)</b>	-0.2622 (0.7960)
Growth then Slump	<b>-3.8238</b> <b>(0.0001)</b>	<b>-2.5814</b> <b>(0.0098)</b>	<b>-7.0386</b> <b>(0.0000)</b>	<b>-2.9391</b> <b>(0.0084)</b>
Growth then Slump, Available Expert Data	<b>-3.8249</b> <b>(0.0001)</b>	-1.0657 (0.2866)	<b>-6.1101</b> <b>(0.0000)</b>	-1.2971 (0.2101)

Note: Statistically significant differences between the human behaviour median and a Bellman-optimal model are marked in bold.

The first important finding is that the theoretically best solution of Eqs. (1)–(2), denoted in Table 1 as Full Information Model, remains unattainable by the averaged human choices. That Bellman-optimal trajectory is superior under all four experimental conditions. Anyway, it was unrealistic to expect anything else, and all *p*-values indicate quite strong differences.

It is remarkable that on average, participants did rather well under three of all four conditions when compared with the Probabilistic Information Model of Eqs. (3)–(4). They managed to adapt to the supplier actions and to develop intuitive statistical understanding of what to anticipate from the latter. Noteworthy is the condition described by production growth *and* available expert data, where the *p*-values suggest almost no distinction between human performance and the probabilistic Bellman-optimal decision trajectory.

A deeper insight may be derived from Fig. 2. Plot a) shows how the median follows the optimal path quite well over the first 10–11 rounds, meaning that each player received efficient support by the others’ messages. A variability increase towards the end of the game might be attributed to the rising omnium bonum quantities.

Somewhat opposite behaviour is observed in Fig. 2b, where the human variability is large early in the game and diminishes in the second half. A likely reason is that people needed more time to adapt to the experiment complexity, which

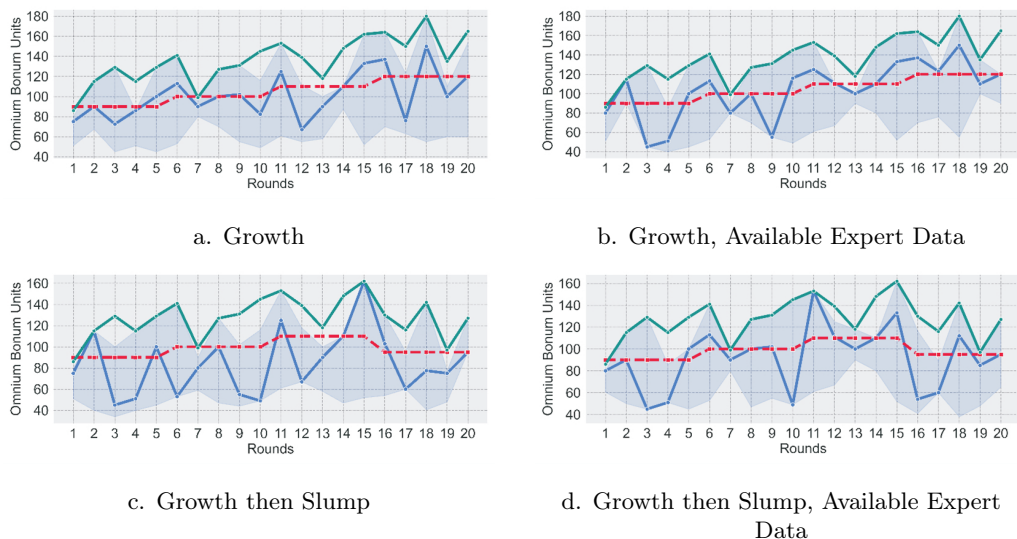


Fig. 2. Bellman-optimal decision trajectories and actual human behaviour. The green line shows the optimal choices under full information; The dotted red line is the optimal sequence under probabilistic information. The blue line and shaded area are the human choices median and an interquartile range

involved additional expert data in the form of two production indicators. Midway through the rounds, people got used to that information and absorbed it.

Where the players could really not compete with the Bellman-optimal models was in the ‘Growth then Slump’ condition 2c. Struggling to understand the game sufficiently in the first dozen rounds, soon they faced a surprising production decrease that could not be dealt with to the very end. It is quite plausible that the common feeling of dissatisfaction, spreading over the virtual social network at that time, contributed to more emotional and less rational actions, leading further away from optimality.

Put in similar circumstances, the players in condition 2d, ‘Growth then Slump, Available Expert Data’ could at least balance the negative emotion with additional expert information. Thus, their median achievement was better than in 2c, and again was statistically not different from the Probabilistic Information Model. That was a pleasant surprise, and perhaps the only unexpected outcome for the researchers.

**Discussion.** These experimental findings lead to some interesting insights. Engineers are groomed in the KEENY and RAIFFA [7] tradition, which not only values rational decisions based on expert information, but also takes them somehow for granted. However, in the same decade (1970s) when their influential book [7] was published, another prominent duo, TVERSKY and KAHNEMAN, showed how and when people deviate substantially from the rationality standard [8,9], and

can fall prey to positive and negative problem formulations within minutes [10].

Our experimental design combined the effects of expert information and emotional reaction on decision making. While the task content stimulated maximal use of the available data, the influence of everybody else in the virtual network provoked more easy-going intuitive choices. That approach worked well in times of production growth, in the sense that people's achievements approached Bellman optimality. During a slump, however, the network fuelled negative sentiment leading to much poorer choices. All of that supports the old finding that a consensus may not be the optimal policy in forecasting and decision making because a common deficiency in reasoning would also lead to a consensus [11].

**Conclusion.** Here we reported the results of an experiment whereby expert intuition was a major factor in choices regarding an abstract production system. At the same time, a virtual social network intimidated the decision-making process by amplifying the predominant dispositions. In contrast, when additional expert information was supplied to the participants, it not only improved their achievements but also served as an antidote for emotional contamination. We can conclude that under all circumstances, it is advisable to search for and rely on high quality data when making important decisions.

## REFERENCES

- [1] HUANG J., X. CHEN, S. MISHRA et al. (2024) Large language models cannot self-correct reasoning yet. In: The Twelfth International Conference on Learning Representations, <https://doi.org/10.48550/arXiv.2310.01798>.
- [2] GRIOT M., C. HEMPTINNE, J. VANDERDONCKT, D. YUKSEL (2025) Large Language Models lack essential metacognition for reliable medical reasoning, *Nat. Commun.*, **16**, 642, <https://doi.org/10.1038/s41467-024-55628-6>.
- [3] REYNA V. F., C. J. BRAINERD (1991) Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion, *J. Behav. Decis. Mak.*, **4**(4), 249–262, <https://doi.org/10.1002/bdm.3960040403>.
- [4] REYNA V. F., C. J. BRAINERD (2008) Numeracy, ratio bias, and denominator neglect in judgments of risk and probability, *Learn. Individ. Differ.*, **18**, 89–107, <https://doi.org/10.1016/j.lindif.2007.03.011>.
- [5] MENGGOV G., N. GEORGIEV, I. ZINOVIEVA, A. GERUNOV (2022) Virtual social networking increases the individual's economic predictability, *J. Behav. Exp. Econ.*, **101**, 101944, <https://doi.org/10.1016/j.socec.2022.101944>.
- [6] GRIFFITHS T. L., J. B. TENENBAUM (2006) Optimal predictions in everyday cognition, *Psychol. Sci.*, **17**(9), 767–773, <https://doi.org/10.1111/j.1467-9280.2006.01780.x>.
- [7] KEENY L., H. RAIFFA (1976) *Decisions and Multiple Objectives*, John Wiley, Hoboken, <https://doi.org/10.1017/CB09781139174084>.

- [8] TVERSKY A., D. KAHNEMAN (1974) Judgement under uncertainty: Heuristics and biases, *Science*, **185**, 1124–1131, <https://doi.org/10.1126/science.185.4157.1124>.
- [9] KAHNEMAN D., A. TVERSKY (1979) Prospect theory: An analysis of decision under risk, *Econometrica*, **47**(2), 263–291, <https://doi.org/10.2307/1914185>.
- [10] KAHNEMAN D., A. TVERSKY (1984) Choices, values and frames, *Am. Psychol.*, **39**(4), 341–350, <https://doi.org/10.1037/0003-066X.39.4.341>.
- [11] KAHNEMAN D., D. LOVALLO (1993) Timid choices and bold forecasts: A cognitive perspective on risk taking, *Manag. Sci.*, **39**(1), 17–31, <https://doi.org/10.1287/mnsc.39.1.17>.

*Faculty of Economics and Business Administration,  
Sofia University “St. Kliment Ohridski”,  
125 Tsarigradsko Shosse Blvd, Bl. 3, 1113 Sofia, Bulgaria  
e-mails: i.atanasov@feb.uni-sofia.bg, g.mengov@feb.uni-sofia.bg,  
A.Gerunov@feb.uni-sofia.bg*