The Effect of Timing When Introducing a Decision Aid in a Decision Support System for Supply Chain Management

Nirit Gavish[(✉)] and Hussein Naseraldin

Ort Braude College, Karmiel, Israel
{Nirit,nhussein}@braude.ac.il

Abstract. In the current research, we evaluated the effect of previous experience with a task on users' willingness to accept a suggested decision aid and their performance in a supply chain management system. Participants were randomly assigned to one of three between-participants groups: the No Aid Group, which did not receive any decision aid during the interaction; the Aid Group, which was offered an algorithm, at the beginning of the first session, to help members improve their decisions; and the Mid-term Aid Group, which received, at the beginning of the second session, the same algorithm that was offered to the Aid Group. The results demonstrated that in the second session the performance of the Aid Group was significantly better in comparison to the No Aid Group. The Mid-term Aid Group's performance did not show any gain, although both the Aid and Mid-term Aid Groups made similar use of the aid in this session. We concluded that in a decision support system, previous experience with the task before introducing the decision aid is not always helpful.

Keywords: Decision aid · Decision support system · Experience · Algorithm · Compliance · Reliance

1 Introduction

Decision support systems (DSSs) in the domain of Industrial Engineering and Management (IE&M) can be very powerful tools, but how much do we know about the way users utilize these systems? Do they exploit the systems' strengths or do they hamper the system from maximizing its effectiveness? More specifically, what are the factors that cause users to adopt the system's decision aids so that they can improve their performance? In the current study we focused on the issue of previous experience: What is the impact that a decision aid has on the performance of a user who has past experience with a task, but without the decision aid?

When developing DSS for IE&M, special care should be given to the user's willingness to use the decision aid. Past studies have shown that this willingness is affected by several factors. For example, Rice and Keller [1] demonstrated that time pressure increased willingness to accept the system's recommendations. Dzindolet et al. [2] showed that information about the possible reasons behind the system's incorrect decisions increased reliance on the system. Lacson et al. [3] showed that

willingness to rely on the system is affected by presenting its reliability in terms of correct or incorrect diagnoses. Systems, however, may be as fallible as humans and users must also exercise judgment when accepting the system's help. They should consider when and how the decision aid helps increase performance levels. Several examples in the literature present both under-reliance on the system's aid, when users ignored the aid that could assist their performance [4, 5, 6], and overreliance, when users accepted bad decisions offered by the system [7, 8]. A successful DSS, accordingly, is one that is both utilized efficiently and improves performance.

In the current study, we focused on the effect on users' willingness to accept the aid when they had previous experience with the task, and on their performance. It is important to examine whether a decision aid should be introduced to the user very early in her interaction with the system or later on. We questioned whether previous training on performing the task, without the decision aid, result in a higher or a lower acceptance level of the decision aid when the decision aid is given, and does it result in better or worse performance measures then?

We evaluated the effect of previous experience on users' willingness to accept the support of a simple decision aid and their performance operating a supply chain management system. Participants were invited to a computer lab and had two successive sessions interacting with a simulation-based supply chain game. The decision aid we adopted in this research is a simple aid that is not automatically produced by the system. We offered participants a simple algorithm and a formula that could help them make a better decision; yet, the aid can shed some light on the way users accept and utilize decision aids in general. Our hypotheses were that since the task is difficult to perform, having had previous experience with the task, users know the task is difficult and hence will be more inclined to accept the decision aid when the decision aid is given. Previous experience should, theoretically, lead to better performance.

2 Method

2.1 Design

Participants were randomly assigned to one of three between-participants groups: the No Aid Group, which did not receive any decision aid during the interaction; the Aid Group, which was offered an algorithm at the beginning of the first session, to help members improve their decision making; and the Mid-term Aid Group, which received, at the beginning of the second session, the same algorithm that was offered to the Aid Group. The randomized assignment was achieved by letting participants choose their preferable date and hour for the experiment, and then assigning each hour randomly to one experimental condition, but controlling for an equal proportion of genders among the groups. The design was mixed, with the session as the within-participants repeated measure independent variable, and the group (No Aid Group, Aid Group and Mid-term Aid Group) as the between-participants independent variable, and two dependent variables: mean total cost and mean deviation from the algorithm.

2.2   Participants

One hundred undergraduate students (66 % males, 34 % females) from ORT Braude College, Israel, participated in the experiment. Participants' average age was 25.2, with a range of 18–51. Ten percent of the participants (19 % in the No Aid Group, 3 % in the in Aid Group, and 9 % in the Mid-term Aid Group) were from the Department of Industrial Engineering and Management at ORT Braude, and the others were from several other engineering departments. Note that in courses given in the Department of Industrial Engineering and Management students are exposed to the topic of supply chain management. It is important that only ten were from this department, and that they were assigned to all experimental groups.

Participants were paid a fixed amount of NIS 40 (about USD 10) for their participation, and given bonuses, which depended on their performance (minimum total cost): The best performer in the respective condition (out of the three groups) received a bonus of NIS 100, and four runner-ups got a bonus of NIS 50 each.

2.3   Apparatus

The experiment took place in a computer lab at the college. Sixteen desktop computers were used, each having a 19-inch monitor. The dedicated program for this experiment was downloaded for each computer; see Experimental Task section. Papers, pencils and calculators were provided to participants.

2.4   Experimental Task

The experimental task was a variation of the Beer Distribution Game, which was developed at MIT in the 1960s and has been widely used to educate graduate students and business managers about supply chain dynamics [9–12]. The version used in the current study was downloaded from the site http://www.runthemodel.com/models/run. php?popup=1&id=507 and was developed by XJ Technologies©, www.anylogic.com; see the experimental task screen layout in Fig. 1. Participants were assigned the role of Retailer, and the computer played the roles of Wholesaler, Distributer, and Factory. For all four roles, the Initial inventory was set at 100.

The experimental task required participants to determine the daily amount to order from their supplier so as to reduce their total costs. The costs for each day included storage costs (for the inventory) and backlog costs (in case a participant could not satisfy customer demand on the same day – negative inventory). Each of the two sessions included 60 days of running the simulation. Participants' goal was to reach the minimum cumulative costs at the end of each session of 60 runs. The inventory was updated every day. Participants could also see, for each day, the following data:

- Expected: The amount ordered and not yet arrived.
- Shipped: The amount shipped the same day to the customer.
- Ordered: The amount that the customer ordered that day. If the amount was in the inventory, it was supplied to the customer—in which case, the shipped amount was
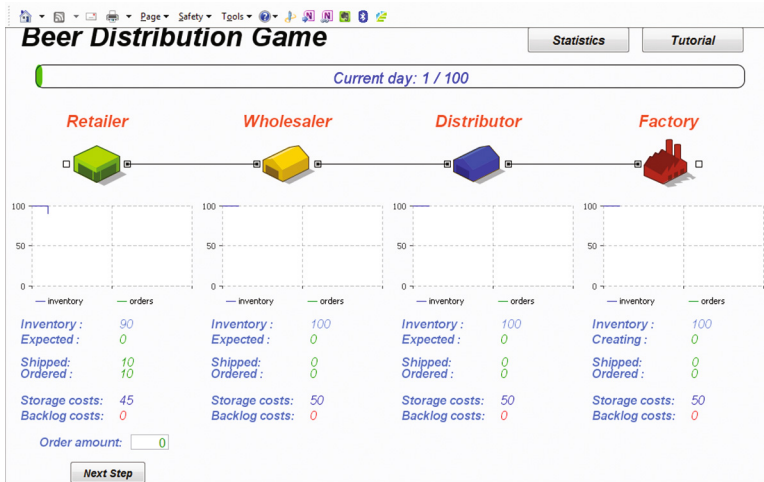
Fig. 1. The experimental task screen layout.

identical to the order amount. If the inventory was smaller than the ordered quantity, the customer was provided with the available inventory and the remainder was provided later when the inventory became available.

The cumulative holding cost was given in the Storage costs. Each unit held in inventory on a particular day cost participants 0.5. For example, if a participant had 3 units in stock on a particular day, that day's storage cost was 1.5. The cumulative cost of holding stock accrued daily. The cumulative backlog cost was given in Backlog costs. This is the cost incurred when a customer ordered merchandise that could not be provided. Each unit missing on a particular day cost 1. For example, if the inventory on a particular day was (−3), the backlog cost for that day was 3. The cumulative backlog cost accrued from day to day. When a certain amount was lacking, participants paid for it until the shortfall was filled. Since the cost of a backlog unit was twice as much as an inventory unit, it was rational to pay more attention to decreasing the backlog than to decreasing the inventory. However, no formal instructions about this were given to participants.

The lead time of each order was four days. For example, if on day 10 participants ordered 20 units, they would receive them on day 14. Once participants decided on the Order amount, they clicked "Next Step" to move to the next day.

The customer orders in each session were, in fact, identical. However, this information was not given to participants. All participants were exposed to the same demand scenarios.

2.5   Procedure

Ten participants at a time sat in the lab but each participant worked individually at his desktop. Each group was assigned randomly to one experimental condition and was blind to the other conditions. On average, the entire experiment took about 45 min and

no more than 75 min. An experimenter remained in the lab, instructed the participants, and presented the algorithm to the relevant groups at the relevant stage.

Participants received a consent form to sign and completed a personal details questionnaire. Following this, the experimenter read aloud an explanation about the experiment and the game. The game, its menus, screens, etc. were explained to participants as they sat in front of their computers. They also received a written manual about the game to assist them during the experiment. Thereafter, participants practiced playing the game for three periods (days) with a self-tutorial, lasted about 5 min. Once this stage was completed, the experimenter confirmed that the participants successfully followed the practice's instructions. Then, participants from the Aid Group received a written explanation about the algorithm, and had it read aloud to them by the experimenter. Participants played the game over one session equaling 60 days, each at his or her individual pace. They were instructed to write down in a table their data and decisions for each day. The data and the decisions (orders) were analyzed. When they finished, participants from the Mid-term Aid Group received the explanation about the algorithm in writing and had it read it aloud to them by the experimenter. Participants from the Aid Group and from the No Aid Group did not receive any additional instructions. Participants again played the game over one session equaling 60 days. When completed, participants were thanked and paid for their participation.

3   Results

Participants' performance and decisions in the first and in the second session were analyzed using a multivariate analysis. In the first session, the results of the No Aid Group and the Mid-term Aid Group were combined (and the combined group was termed "No Aid - First Session Group"), because at this stage of the experiment the manipulation they were exposed to was identical. Two measures were analyzed for evaluating the performance and acceptance level. These measures were mean total cost (sum of storage and backlog costs for each day) and the mean deviation from the algorithm (measured as the absolute value of the order amount recommended by the algorithm minus the participant's order amount for each day). Although the mean deviation from the algorithm was not a perfect indicator of the acceptance level, it reflected how much the participant's decisions were close to the suggestion of the algorithm. As was demonstrated, this later measure did not have to be connected to better performance since the algorithm was not ideal to achieve the best score possible.

3.1   First Session Analysis

In the multivariate analysis, the group (No Aid - First Session Group and Aid Group) was the independent variable; the mean total cost and the mean deviation from the algorithm were the dependent variables.

The multivariate analysis demonstrated that the effect of group was not significant (Wilks' Lambda test on the combined variable: ($F(2,97) = 2.5$, $p = 0.08$, *Partial Eta Squared* = 0.05). A positive significant correlation was found between the two dependent variables (*Pearson r* = 0.6, $p < 0.001$). A univariate analysis was performed

for each of these variables, and demonstrated that for the mean total costs, the effect of group was not significant ($F(1,98) = 3.6$, $p = 0.06$, *Partial Eta Squared* = 0.04). For the mean deviation from the algorithm, the effect of group was significant ($F(1,98) = 4.4$, $p = 0.04$, *Partial Eta Squared* = 0.04), with higher deviation from the algorithm to the No Aid – First Session Group ($M = 30.0$, $SD = 46.1$) compared to the Aid Group ($M = 12.5$, $SD = 17.9$).

3.2   Second Session Analysis

The multivariate analysis demonstrated that the effect of group was significant (Wilks' Lambda test on the combined variable: ($F(4,192) = 4.3$, $p = 0.002$, *Partial Eta Squared* = 0.08). A positive significant correlation was found between the two dependent variables (*Pearson r* = 0.7, $p < 0.001$). Reported below are the univariate results of these variables.

For the mean total costs, the effect of group was significant ($F(2,97) = 4.0$, $p = 0.021$, *Partial Eta Squared* = 0.08). Post hoc Tukey HSD test showed that No Aid Group had a significantly higher mean total cost ($M = 45.5$, $SD = 27.7$) compared to the Aid Group ($M = 26.1$, $SD = 27.7$; $p = 0.015$), while the other contrasts were not significant (No Aid Group compared to the Mid-term Aid Group ($M = 34.3$, $SD = 27.7$): $p = 0.2$; Aid Group compared to the Mid-term Aid Group: $p = 04$).

For the mean deviation from the algorithm, the effect of group was significant ($F(2,97) = 6.1$, $p = 0.003$, *Partial Eta Squared* = 0.1). Post hoc Tukey HSD test showed that No Aid Group had a significantly higher mean deviation from the algorithm ($M = 20.3$, $SD = 15.2$) compared to both the Aid Group ($M = 10.0$, $SD = 15.2$; $p = 0.02$) and the Mid-term Aid Group ($M = 8.1$, $SD = 15.2$; $p = 0.004$). The difference between the Aid Group and the Mid-term Aid Group was not significantly different ($p = 0.9$).

4   Discussion

The results demonstrated that generally, when exposed to the algorithm, participants tended to use it. In the first session, the Aid Group's decisions were closer to the algorithm than these of the No Aid - First Session Group; in the second session, there was a significant difference in the deviation from the algorithm between the two groups (Aid Group and Mid-term Aid Group) in comparison to the No Aid Group. In addition, it seems that using the algorithm was intuitive and participants did not need to gain experience in learning how to use it, as both the Mid-term Aid Group, which was exposed to the algorithm in the second session, used it in that session no less often than did the Aid Group, which had previous experience with it.

The Aid Group had significantly better performance compared to the No Aid Group in the second session, while no significant difference was demonstrated between the Mid-term Aid Group and the No Aid Group. In contrast, the degree of utilization of the decision aid was similar in both groups that were exposed to it in the second session. Only the group that received the algorithm at the beginning of its first session of interaction with the system improved its performance better. The contrast between the acceptance level and performance results is especially interesting since the two

measures were correlated: Using the algorithm indeed can lead to better performance, but it seems that mere using it is not enough, and it should be done carefully.

Our hypotheses that previous experience with the task without the decision aid will result in higher acceptance level of using the decision aid and lower total costs compared to experience with the decision aid only were not confirmed. The two groups that received the aid demonstrated a similar acceptance level, and the group that had previous experience with the aid performed better. We conjectured that the previous experience with the decision aid did not change the mean acceptance levels, but did change their nature: Since the decision aid was not optimal, participants learned how to utilize it better, to use it more when it is constructive and less when it led them to poorer performance.

The conclusions from the current study are that in a decision support system, previous experience performing a task without decision aids is not only not necessary, but can also delay the process of learning how to best utilize the decision aids—most importantly, avoiding both under-reliance [4–6] and overreliance [7, 8].

References

1. Rice, R., Keller, D.: Automation reliance under time pressure. Cogn. Technol. **14**, 36–44 (2009)
2. Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., Pierce, L.G., Beck, H.P.: The role of trust in automation reliance. Int. J. Hum. Comput. Stud. **58**, 697–718 (2003)
3. Lacson, F.C., Wiegmann, D.A., Madhavan, P.: Effects of attribute and goal framing on automation reliance and compliance. In: Proceedings of the Human Factors and Ergonomics Society 49 h Annual Meeting, pp. 482–486. Human Factors and Ergonomics Society, Santa Monica (2005)
4. Dzindolet, M.T., Pierce, L.G., Beck, H.P., Dawe, L.A.: The perceived utility of human and automated aids in a visual detection task. Hum. Factors **44**, 79–94 (2002)
5. Parasuraman, R., Riley, V.: Human and automation: use, misuse, disuse. Abuse. Hum. Factors **39**, 230–253 (1997)
6. Riley, V.: Operator reliance on automation: theory and data. In: Parasuraman, R., Moulouas, M. (eds.) Automation and Human Performance: Theory and Applications, pp. 19–35. Lawrence Erlbaum, Mahwah (1997)
7. Layton, C., Smith, P.J., McCoy, C.E.: Design of cooperative problem-solving system for en-route flight planning: an empirical evaluation. Hum. Factors **36**, 94–112 (1994)
8. Parasuraman, R., Molloy, R., Singh, I.L.: Performance consequences of automation-induced complacency. Int. J. Aviat. Psychol. **3**, 1–23 (1993)
9. Jackson, G.C., Taylor, J.C.: Administering the MIT beer game: lessons learned. Dev. Bus. Simul. Exp. Learn. **25**, 208–214 (1993)
10. Liu, H., Howley, E., Duggan, J.: Optimisation of the beer distribution game with complex customer demand patterns. In: IEEE Congress on Evolutionary Computation, 2009, CEC 2009, pp. 2638–2645. IEEE (2009)

11. Ravid, G., Rafaeli, S.: Multi player, Internet and java-based simulation games: learning and research in implementing a computerized version of the "beer-distribution supply chain game". Simul. Ser. **32**, 15–22 (2000)
12. Sterman, J.: Instructions for Running the Beer Distribution Game. Systems Dynamics Group. Sloan School of Management, Cambridge, Massachusetts 02139 (1984)