Research Article

An experience-based contrast effect when relying on a decision aid

Nirit Yuviler-Gavish[1] · Hussein Naseraldin[1]

Abstract

The present study investigates experience-based contrast effect as a factor that affects users' decision to rely on decision support systems or not. Two types of experience-based contrast effects were evaluated: "hands-on" and "hard first, simple second." In the first study, testing "hands-on" contrast, we examined its effect by giving users experience with a task before introducing them to a hard-to-use decision aid to carry out the same task. In the second study, examining "hard first, simple second" contrast, the effect was evaluated by introducing users first to a hard-to-use decision aid and thereafter to a simple decision aid. In both studies, participants took part in two successive sessions in a simulation-based supply chain game. Results demonstrated that both types of contrasts reduced participants' reliance on the decision aids. Hence, when encouraging users to rely on a decision aid, one should consider the experience-based contrast effects and use them carefully. The implications from this work can contribute to our understanding of the role of experience-based contrast effect when relying on a decision aid.

1 Introduction

In recent decades, many fields have seen the development of decision support systems (DSSs). These systems equip human decision makers with knowledge and tools to help make reasoned decisions that in the past were based almost solely on intuition and cumulative experience. DSSs are commonly used for forecasting in a wide range of situations that require long-term planning, from finance to climate science [1–6]. Other areas where DSSs have been applied include road safety, e.g., decision support for drivers at intersections [7, 8]; health care, e.g., systems that use patient data to generate case-specific advice for physicians [9–12]; and business strategy [13, 14].

DSS users, who can potentially improve their decision making with this powerful tool, are also prone to two phenomena that hamper them from using the DSS effectively.

These are overreliance, when users accept bad decisions offered by the system [15, 16], and under-reliance, when users fail to exploit the system's capabilities [17–19]. The present paper is concerned with the latter situation.

Relying on the aid of a DSS is influenced by past experience with the DSS's effectiveness and the way it contributes to the decision making process. The DSS's efficiency (whether it leads to faster or effortless decision making) and accuracy (whether its aids are helpful) are factors considered by the users deciding whether to accept the decision aid [20]. Usually, reliance on the DSS develops concurrently with ongoing experience with the system and exposure to its performance [21–27].

The performance aspect of the DSS, however, is apparently not the only parameter that is important to users when deciding to rely (or not) on the aid. Many other factors—not linked directly to the DSS's actual efficiency and

✉ Nirit Yuviler-Gavish, nirit@braude.ac.il; Hussein Naseraldin, nhussein@braude.ac.il | [1]Department of Industrial Engineering and Management, ORT Braude College, Snunit 51, P.O. Box 78, 2161002 Karmiel, Israel.

effectiveness—can affect users' acceptance level of the decision aid. For example, Spain and Madhavan [28] demonstrated that participants tended to trust and comply less with a decision aid given in a rude manner than when the same decision aid is given in a natural and polite manner. Lacson et al. [29] showed that in an aided signal detection task, framing the system's reliability in terms of 80% correct versus 20% incorrect led to higher reliance among the "80% correct" group. In addition, the same advice of an expert system is perceived differently if it is given in full sentences compared to being given in a production rule style [30]. Also, interestingly, background information about an emergent decision support system, including its functions and the fact that it was a recently developed system whose credibility had yet to be established, led participants to increase their trust in and utilization of the system [31].

We contribute to the above literature by studying the influence of experience-based *contrast effect*s on users' reliance on a decision aid in a DSS. The contrast effect occurs when the person's judgment about someone or something is influenced by his prior exposure to something, which can serve as a benchmark to the to-be-judged object. The contrast effect is well known in the literature. For example, animals were judged as less ferocious or less large when participants were first presented with extremely ferocious or extremely large animals [32]. Women that were exposed to photographs of idealized thin physique models decreased their self-esteem and increased their self-consciousness, social physique anxiety and body dissatisfaction [33]. The contrast effect was even demonstrated in newborns, who sucked less water when they were given 15% sucrose before given the water compared to a group that were given only water [34]. Other examples can be found in Chen and He [35]; Dawes et al. [36]; Di Lollo and Beez [37]; Oikawa et al. [38]; Schuh [39]; Smither et al. [40]; and many more.

However, not many studies focused on the contrast effect when relying on a decision aid in a DSS. Rice et al. [41] demonstrated that an exposure to positive or negative exemplars of automation affected trust in a decision aid and the actual use in it. Yang et al. [42] showed that the human operator evaluates an automated technology by benchmarking it against his own ability, rather than against predetermined objective criteria.

In this paper, we contributed to the above literature by examining experience-based contrast effects on the users' reliance on a DSS. We performed two studies to evaluate how different types of experience-based contrasts, termed "hands-on" and "hard first, simple second," affect the reliance on a decision aid given in a supply chain management setting. In the first study, a "hands-on" contrast effect was evaluated. Users acquired experience with a task

before being introduced to a hard-to-use decision aid. In the second study, a "hard first, easy second" contrast effect was evaluated. We presented a hard-to-use decision aid (identical to the one used in the first study) before introducing a simple decision aid.

Our hypotheses regarding the "hands-on" versus the "hard first, easy second" contrast effects were that the first would decrease reliance on the decision aid, and the second would increase it. For the "hands-on" contrast effect, we hypothesized that previous experience with the task without a decision aid would convince participants that the task can be done without the aid, and hence, they would be less inclined to rely on the decision aid we offered, especially when it is hard to use. (In Yuviler-Gavish and Naseraldin's study [43], using a simple decision aid instead of an hard to use one did not make participants with previous task experience relying on the aid less.) This hypothesis is based on both Yang et al's [42] finding about the way users benchmark the decision aid against their own ability and several studies which demonstrate that users with grater subject matters expertise are less likely to rely on automation that novice users are [44–46]. Although participants will not become expert after such a short period of interaction with the task, their benchmark will change. For the "hard first, easy second" contrast effect, we hypothesized, based on the many studies on the contrast effect, that the reliance on a simple decision aid will be higher compared to no benchmark condition, since participants would perceive their investment in adopting the simple aid as lower if they had already experienced working with a more demanding decision aid.

Two pilot studies were performed, one for each contrast effect, to ensure that the system works well, participants understand the task and the instructions, and that the duration of the experiment is no more than 1 h. The protocols were amended according to the pilots' results.

2 Study 1: "Hands-on" contrast effect

2.1 Method

This study used the same simulation-based supply chain management system used in Yuviler-Gavish and Naseraldin's [43] work and a very similar method.

2.1.1 Design

Seventy-three students from ORT Braude College took part in the study. Participants were invited to a computer laboratory for two successive sessions with a simulation-based supply chain game. In the first session, 27 of the 73 participants were given an algorithm at the start of the

first session and informed that using this algorithm could help them improve their decision making (Aid group). Twenty-six of the remaining 46 participants were given the algorithm at the start of the second session (Mid-term Aid group), and the remaining 20 participants continued with no support (No Aid group). Participants were randomly assigned to each of the three groups, ensuring that the total number of participants in each group will be 20 or more, and that the percentages of males and females will be similar among the groups. The design is illustrated in Table 1.

2.1.2 Participants

All 73 participants were undergraduate students from ORT Braude College, Israel (65% males, 35% females: 56% males in the Aid group, 69% males in the Mid-term Aid group and 75% males in the No Aid group). All participants were engineering students, to ensure that they have the mathematical understanding needed for this task, to decrease the variability in the results, and because the task was from the type of engineering problems that they might face with during their future careers. Participants' average age was 25.6, with a range of 20–35. None of the participants had participated in Yuviler-Gavish and Naseraldin's [43] study. Participants were paid a fixed amount of NIS 40 (about USD 10) for their participation, along with bonuses based on their performance in the game (see below). The best performer in each condition received a bonus of NIS 100, and four runners-up in each group received a bonus of NIS 50 each. This research complied with the American Psychological Association Code of Ethics and was approved by the Ethical Committee at ORT Braude College. Informed consent was obtained from each participant.

2.1.3 Apparatus and procedure

The experiment took place in a computer laboratory at the college; each computer had a 19-in. monitor. Participants were invited in groups of up to 10, but worked individually at their own desktops. The dedicated program for this experiment was downloaded to each computer (see under "3.1.4"). Participants were given paper, pencils and calculators to use during the experiment.

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 laboratory throughout the experiment, gave participants the appropriate instructions and presented the algorithm to the relevant groups at the relevant stage.

Participants first signed a consent form and completed a personal details questionnaire. Following this, the experimenter read aloud the instructions for the game, explained the various menus and screens that participants would encounter, and showed participants how to input their data and decisions into a table. Participants were also given a written manual containing the game instructions. Participants then used a self-tutorial incorporated in the game program to practice playing the game for three periods ("days"; see "3.1.4"). Following this exercise, which took about 5 min, the experimenter confirmed that all participants understood how to play the game.

At this stage, in the No Aid condition, participants played the game at their own pace for one session covering 30 periods ("days"), recording their data and decisions ("orders"; see below) for each day as instructed. After a short break, participants again played the game for a second session, this one totaling 30 periods (days). At that point, participants were thanked and paid for their participation.

The procedure for the Aid and Mid-term Aid groups was the same, with the following exceptions. For the Aid group, the decision support algorithm was distributed and read aloud by the experimenter after the self-tutorial and confirmation by the experimenter that participants understood the game (i.e., at the start of Session 1). For the Mid-term Aid group, the decision support algorithm was distributed and read aloud at the start of Session 2. The experimenter made no mention of the algorithm to any group at any point other than upon first introducing it where relevant. Participants were not given practice sessions with the algorithm, but were told that using it would be the most effective way to improve their scores (see below).

Participants' decisions and performance during the second session were recorded and analyzed.

2.1.4 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 [47–50]. The version used in the current study was developed by XJ Technologies©, www.anylogic.com, and was downloaded from the site https://cloud.anylogic.com/model/b0156

Table 1 Study 1: Design—X denotes that the aid was given

	No Aid group	Mid-term Aid group	Aid group
First session			X
Second session		X	X

f6d-6c04-431b-b48d-1b875b2720e7?mode=SETTINGS
[51]. Figure 1 presents the experimental task screen layout, and Fig. 2 shows the screen layout with some of the explanations seen by participants during the instruction stage. 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 units.

The experimental task required participants to determine the number of units (e.g., cases) of beer to order from their supplier each day so as to fulfill all orders from customers while minimizing their cumulative costs at the end of each session (30 "days" for each of the sessions). Costs included each day's accumulated inventory storage costs and backlog costs (the value of any orders from customers which the participant was unable to supply—i.e., negative inventory).

Play proceeded as follows: each "day," the "retailer" (the participant) received a given order from a "customer" (i.e., the computer program). The participant was expected to fulfill ("ship") that order immediately to the extent that the number of units required were in storage; any remaining units would have to be ordered from the wholesaler. The lead time for each order was 4 days, such that, for example, a participant who ordered 20 units on day 10 would receive those units on day 14. Participants had to decide how much to order from the wholesaler on any given day based on previous and anticipated orders from customers,

while taking into account the costs associated with storage and backlogs (on which see below). Selecting an amount to order each day ended that period's play; participants then clicked "Next Step" to move to the next day. The inventory was updated every day. Participants could also see, for each day, the number of units ordered by the retailer from the wholesaler and not yet arrived; the number of units ordered by the customer; and the number of units shipped to the customer.

The cost of holding inventory (storage) was set at 0.5 units of currency per unit of stock per day. For instance, storing three units of stock for 1 day would entail a storage cost of 1.5. Backlog costs were incurred when a customer ordered merchandise that could not be provided, and were set at 1 unit of currency per undelivered unit per day. Thus, a backlog of three units (an inventory of − 3) over one day would entail a backlog cost of 3. The storage and backlog costs accrued from day to day.

Since a backlog unit cost is twice as much as an inventory unit, it was rational for players to prioritize reducing their backlog over reducing their inventory. However, no formal instructions about this were given to participants.

The program was set so that all participants were exposed to the same demand scenarios. That is, each participant received identical customer orders, though participants were not aware of this. The program was further set so that the optimal total cost at the end of each session was 177 (based on a cumulative storage cost of
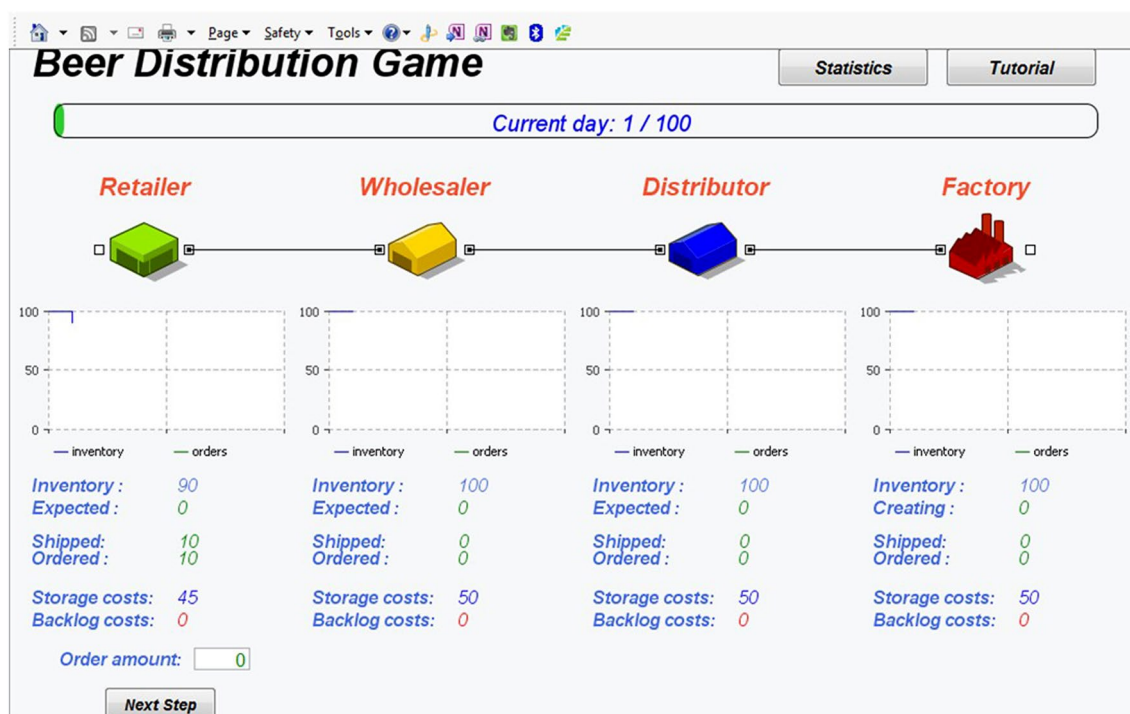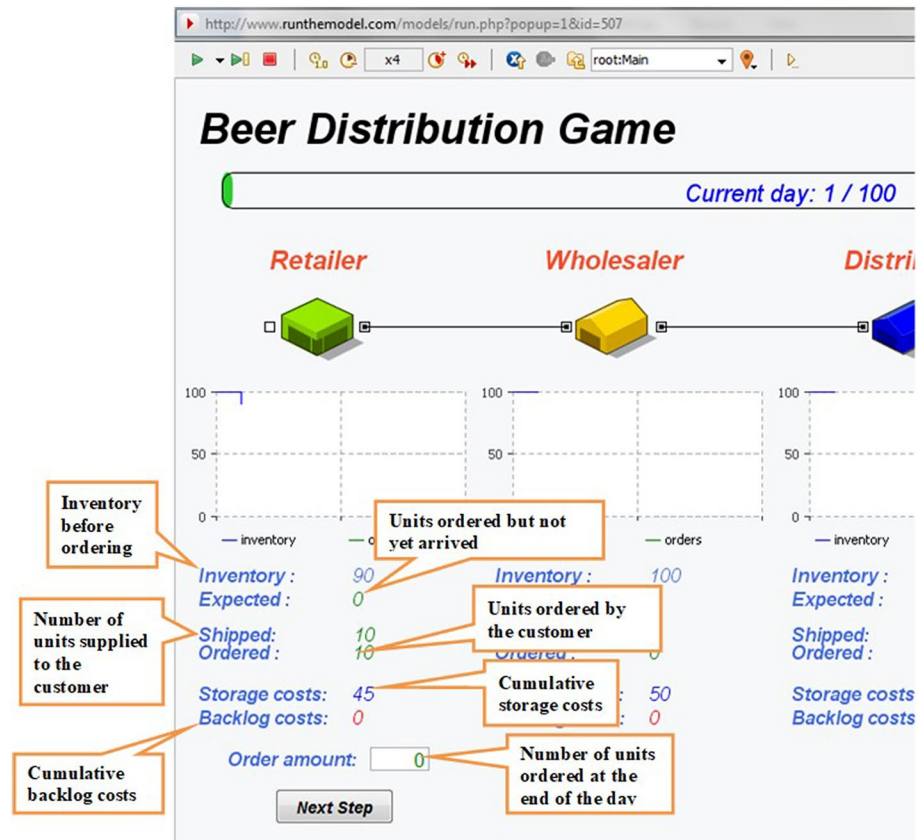

Fig. 1 Experimental task screen layout (taken from www.anylogic.com)

Fig. 2 Experimental task screen layout (taken from www.anylogic.com) with explanations (as seen by participants)

177 and a cumulative backlog cost of 0, when getting rid of the initial inventory and then ordering exactly the needed amount for each day), and the optimal mean total cost per day was 5.9. The worst possible total cost was unlimited, since there was no limit to how much participants could order from the wholesaler.

Given perfect knowledge of customers' future orders, players could in theory optimize their total costs by keeping their inventory at 0 and placing each day's orders so as to ensure an accurate stock count 4 days ahead (the lead time for each order). However, in the absence of such perfect knowledge, the algorithm supplied to the Aid group and the Mid-term Aid group provided a way to calculate orders so as to consistently and effectively reduce total costs.

The algorithm and the preamble ran as follows.

Below is an algorithm that will help you calculate order quantities so as to most effectively reduce your costs.

The recommendation is to place orders according to the following policies:

On Day t:

1. *Calculate the forecast for the next day F_{t+1} by a formula that takes into account the forecast for the current day F_t and the customer's order the same day D_t, and rounds*

the result upwards. The forecast for day $t+1$ will be calculated by:

$$F_{t+1} = \lceil 0.9 * D_t + 0.1 * F_t \rceil$$

where the forecast for the first day is 10:

$$F_1 = 10.$$

2. *After you have calculated this, the amount you should order at the end of day t is calculated by a formula that takes into account the current day's forecast, F_t, and the next day's forecast, F_{t+1}, and rounds the result upwards. That is, the amount you should order at the end of day t will be calculated by:*

$$Q_t = \lceil D_t + 4 * (F_{t+1} - F_t) \rceil$$

Example The customer's order on day 1 is 10. The forecast for day 2 is calculated by:

$$F_2 = \lceil 0.9 * D_1 + 0.1 * F \rceil_1 = \lceil 0.9 * 10 + 0.1 * 10 \rceil = \lceil 9 + 1 \rceil = 10$$

The amount to order at the end of day 1 is calculated by:

$$Q_1 = \lceil D_1 + 4 * (F_2 - F_1) \rceil = \lceil 10 + 4 * (10 - 10) \rceil = 10$$

The customer's order on day 2 is 11. The forecast for day 3 is calculated by:

$$F_3 = \lceil 0.9 * D_2 + 0.1 * F \rceil_2 = \lceil 0.9 * 11 + 0.1 * 10 \rceil = \lceil 9.9 + 1 \rceil = \lceil 10.9 \rceil = 11$$

The amount to order at the end of day 2 is calculated by:

$$Q_2 = \lceil D_2 + 4 * (F_3 - F_2) \rceil = \lceil 11 + 4 * (11 - 10) \rceil = 15$$

Based on the customer's orders, Q_t was always greater than 0. Perfect use of the algorithm would produce a total cost of 977.5 for each session (cumulative storage cost: 977.5; cumulative backlog cost: 0), with a mean total cost per day of 32.6. (Recall that the optimal total cost, based on perfect knowledge, was 177, with an optimal mean total cost per day of 5.9.)

2.2 Results

Participants' performance and decisions in the second session were analyzed using multivariate analysis. Two measures were analyzed: mean total costs per day (i.e., the sum of storage and backlog costs for each day) and mean deviation from the algorithm (measured as the absolute difference between the order quantity recommended by the algorithm and the order quantity placed by the participant's, for each day). The first measure, mean total costs per day, was used to evaluate participants' performance. The second, mean deviation from the algorithm, was used to measure acceptance of the decision aid, although it is not a perfect indication of using the algorithm, because more advanced methods to evaluate this (e.g., artificial intelligence) were not available, and simply asking participants is not reliable enough. The full set of results is given in Table 2.

A multivariate analysis was performed, with the group (Aid, Mid-term Aid and No Aid) as the independent variable; mean total costs per day and mean deviation from the algorithm were the dependent variables.

The multivariate analysis revealed a significant effect of group [Wilks' Lambda test on the combined variable: $F(4,138) = 3.0$, $p = 0.02$, partial eta squared $= 0.08$]. A positive significant correlation was found between the two dependent variables (Pearson $r = 0.3$, $p = 0.03$).

We performed a univariate analysis for each variable separately. For mean total costs, the effect of group was not significant [$F(2,70) = 2.8$, $p = 0.07$, partial eta squared $= 0.08$]. These results are shown in Fig. 3.

For mean deviation from the algorithm, the effect of group was significant [$F(2,70) = 4.6$, $p = 0.014$, partial eta squared $= 0.1$]. A post hoc Tukey HSD test showed that the No Aid group deviated significantly more from the algorithm ($M = 14.0$, SD $= 13.2$) than the Aid group ($M = 4.2$, SD $= 2.7$; $p = 0.01$). The other contrasts were not significant: The No Aid group did not deviate significantly more from the algorithm than the Mid-term Aid group ($M = 8.5$, SD $= 14.1$; $p = 0.21$), and the latter did not deviate significantly more than the Aid group ($p = 0.3$). In other words, only participants who had been given the algorithm in Session 1 used it significantly more compared to participants who had not been given it at all. For participants in the Mid-term Aid group, who received the algorithm at the
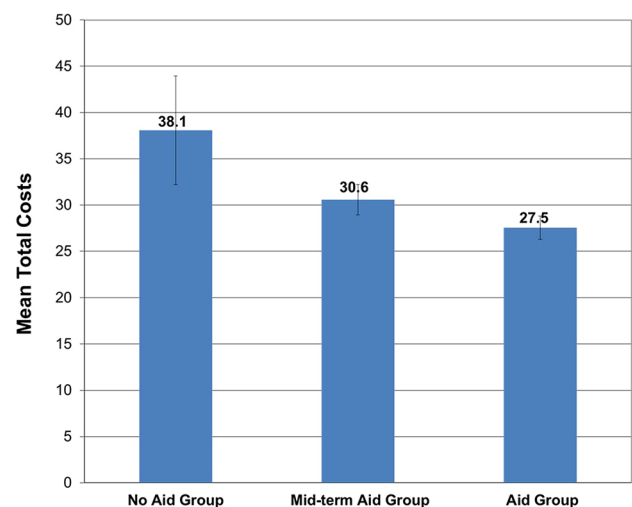

Fig. 3 Study 1: Second session results: Mean total costs for each group (with standard error bars)

Table 2 Study 1: Detailed results for each group, second session—means and standard deviations (in brackets)

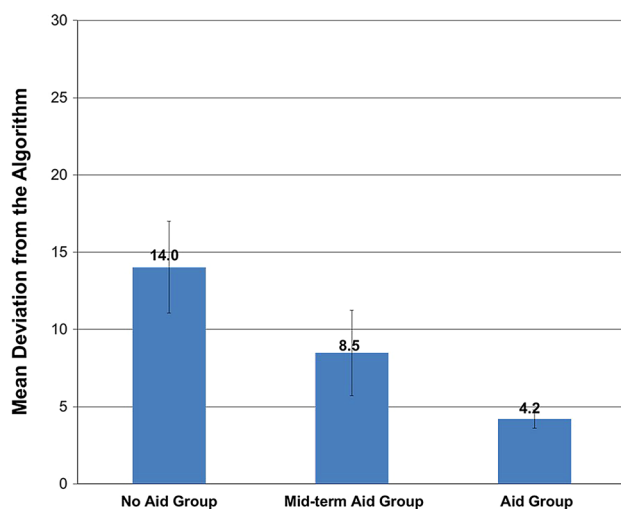| Variable	No Aid group	Mid-term Aid group	Aid group	Optimal policy	Expected with perfect use of the algorithm
Mean total costs	38.1 (26.2)	30.6 (8.4)	27.6 (7.0)	5.9	32.6
Mean deviation from the algorithm	14.0 (13.2)	8.5 (14.1)	4.2 (2.7)		0

Fig. 4 Study 1: Second session results: Mean deviation from the algorithm for each group (with standard error bars)

start of Session 2, the difference was not significant. The results are shown in Fig. 4.

3 Study 2: "Hard first, simple second" contrast effect

3.1 Method

3.1.1 Design

Ninety-one students from ORT Braude College took part in the study. Participants were invited to a computer laboratory for two successive sessions with a simulation-based supply chain game. In the first session, 31 participants were given a hard-to-use algorithm at the start of the first session and were informed that using this algorithm could help them improve their decision making and given a simple algorithm at the start of the second session (Hard-Simple Aid group). Thirty-five participants were given only the simple algorithm at the start of the second session (Simple Aid group). Twenty-five participants were not given any support at all during the two sessions (No Aid group). Participants were randomly assigned to each of the three groups, ensuring that the total number of participants in each group will be 20 or more, and that the percentages of males and females will be similar among the groups. The design is illustrated in Table 3.

3.1.2 Participants

The 91 participants were undergraduate students from ORT Braude College, Israel (70% males, 30% females: 74%

Table 3 Study 2: Design—X denotes that the aid was given

	Aid	No Aid group	Simple Aid group	Hard-Simple Aid group
First session	Simple			
	Hard to use			X
Second session	Simple		X	X
	Hard to use			

males in the Hard-Simple Aid group, 66% males in the Simple Aid group and 72% males in the No Aid group). Participants' average age was 25.6, with a range of 20 to 35. None of the participants had participated in Yuviler-Gavish and Naseraldin's [43] study or in Study 1. Payment conditions were similar to those described in Study 1. This research complied with the American Psychological Association Code of Ethics and was approved by the Ethical Committee at ORT Braude College. Informed consent was obtained from each participant.

3.1.3 Apparatus and procedure

The apparatus and procedure were similar to those described in Study 1.

3.1.4 Experimental task

The experimental task was similar to the experimental task in Study 1.

The simple algorithm and the preamble ran as follows:

Below is an algorithm that will help you calculate order quantities so as to most effectively reduce your costs.

The recommendation is to place orders according to the following policies:

On Days 1, 2, 3 and 4: Order nothing.

On Days 5, 6 and on: Order the average of the customer orders over the past 3 days, rounded up. That is, the amount that will be ordered on day t will be calculated by:

$$Q_t = \left\lceil \frac{D_t + D_{t-1} + D_{t-2}}{3} \right\rceil$$

Example Suppose on days 6, 7 and 8 customer orders are: 12, 14 and 17, respectively; then the amount of the order will be calculated at the end of day 8 by:

$$Q_8 = \left\lceil \frac{D_8 + D_7 + D_6}{3} \right\rceil = \left\lceil \frac{17 + 14 + 12}{3} \right\rceil = 15$$

This algorithm was considered simple because it contained only one calculation stage instead of two, although

the actual calculations made by participants were not observed. The hard-to-use algorithm was similar to the one described in Study 1.

Perfect use of the simple algorithm would produce a total cost of 869.5 for each session (cumulative storage cost: 869.5; cumulative backlog cost: 0), with a mean total cost per day of 29.0.

3.2 Results

Participants' performance and decisions in the second session were analyzed using multivariate analysis with the same two measures described in Study 1. The full set of results is displayed in Table 4.

We performed a multivariate analysis with the group (Hard-Simple Aid, No Aid and Simple Aid) was the independent variable; mean total costs per day and mean deviation from the simple algorithm were the dependent variables.

The multivariate analysis revealed a significant effect of group [Wilks' Lambda test on the combined variable: $F(4,174) = 2.69$, $p = 0.03$, *partial eta squared* $= 0.06$]. A positive significant correlation was found between the two dependent variables (*Pearson r* $= 0.37$, $p < 0.001$).

A univariate analysis was conducted for each variable separately. For the mean total costs, the effect of group was significant [$F(2,88) = 3.70$, $p = 0.03$, *partial eta squared* $= 0.08$]. A post hoc LSD test showed that the No Aid group had significantly higher mean total costs ($M = 39.6$, SD $= 23.3$) compared to both the Simple Aid ($M = 29.8$, SD $= 9.0$; $p = 0.02$) and the Hard-Simple Aid groups ($M = 29.8$, SD $= 13.0$; $p = 0.02$). The contrast between the Simple Aid and the Hard-Simple Aid groups was not significant ($p = 0.99$). In other words, participants who got the simple algorithm were able to reduce their total costs regardless of whether or not they got it after prior exposure to the hard-to-use algorithm. These results are shown in Fig. 5.
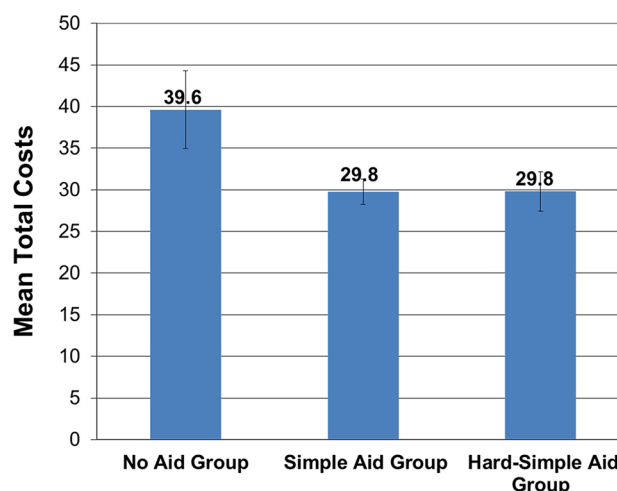

Fig. 5 Study 2: Second session results: Mean total costs for each group (with standard error bars)

For the mean deviation from the simple algorithm, the effect of group was only close to significant [$F(2,88) = 3.70$, $p = 0.06$, *partial eta squared* $= 0.06$]. A post hoc LSD test, however, showed that the No Aid group deviated significantly more from the simple algorithm ($M = 19.3$, SD $= 28.4$) than the Simple Aid group ($M = 4.4$, SD $= 5.6$; $p = 0.02$). The other contrasts were not significant: The No Aid group did not deviate significantly more from the simple algorithm than the Hard-Simple Aid group ($M = 13.5$, SD $= 32.3$; $p = 0.38$), and the latter did not deviate significantly more than the Simple Aid group ($p = 0.13$). In other words, participants who had been given only the simple algorithm in Session 2 relied on it significantly more compared to participants who had not been given it at all. For participants in the Hard-Simple Aid group, who also received, in addition to the simple algorithm in Session 2, the hard-to-use algorithm at the beginning of Session 1, the difference was not significant. The results are shown in Fig. 6.

To examine whether participants in the Hard-Simple Aid group were less willing to rely on the simple algorithm,

Table 4 Study 2: Detailed results for each group, second session—means and standard deviations (in brackets)

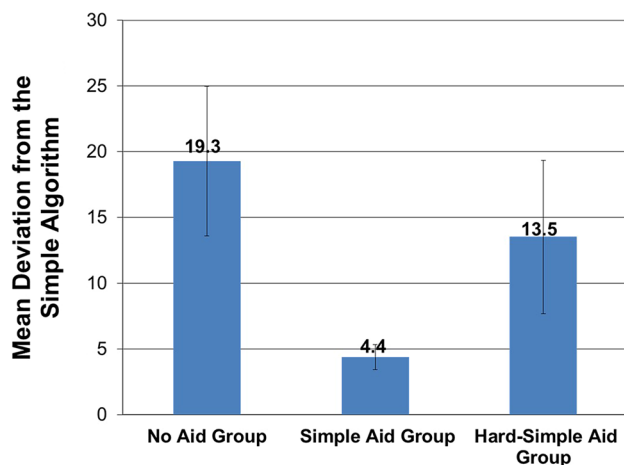| Variable	No Aid group	Simple Aid group	Hard-Simple Aid group	Optimal policy	Expected with perfect use of the hard-to-use algorithm	Expected with perfect use of the simple algorithm
Mean total costs	39.6 (23.3)	29.8 (9.0)	29.8 (13.0)	5.9	32.6	29.0
Mean deviation from the hard-to-use algorithm	35.8 (25.9)		15.8 (32.2)		0	
Mean deviation from the simple algorithm	19.3 (28.4)	4.4 (5.6)	13.5 (32.3)			0

Fig. 6 Study 2: Second session results: Mean deviation from the algorithm for each group (with standard error bars)

in opposition to our hypothesis, because they still relied on the hard-to-use algorithm in the second session, we compared the mean deviation from the hard-to-use algorithm in Session 2 between the Hard-Simple Aid group and the No Aid group. The univariate analysis showed that the effect of group was significant [$F(1,54) = 6.34$, $p = 0.06$, *partial eta squared* $= 0.10$]. Mean deviation from the hard-to-use algorithm was greater for the No Aid group ($M = 35.8$, SD $= 25.9$) compared to the Hard-Simple Aid group ($M = 15.8$, SD $= 32.2$), indicating that participants in the Hard-Simple Aid group were relying on the hard-to-use algorithm in the second session also.

4 Discussion and conclusions

The current work evaluated two kinds of experience-based contrast effects when relying on a decision aid. In the first study, a "hands-on" contrast effect was manipulated via previous experience with a task before introducing the users to a hard-to-use decision aid, and we hypothesized that participants who had previous experience tackling the task without any support would make less use of a fairly hard-to-use decision aid that was later offered to them. In the second study, a "hard first, simple second" contrast effect was manipulated via presenting a hard-to-use decision aid (identical to the one in the first study) before introducing a simple decision aid, and we hypothesized that participants would perceive the investment of time and effort in using the simple decision aid as lower in the presence of prior experience with a hard-to-use decision aid, and hence would rely more on the simple decision aid.

In the first study, the results of the second session demonstrated a significant difference in relying on the decision aid between the Aid group and the No Aid group, while the difference was not evident between the No Aid group and the Mid-term Aid group, suggesting that the Mid-term Aid group did not rely on the decision aid. This was despite the way the algorithm was presented to them, which should have conveyed to participants that they would perform better using the algorithm in the second session than they had in the first. The results thus validate our assumption that the "hands-on" contrast effect reduces people's reliance on a decision aid if it is introduced later (after the task has already been carried out without it), at least when the aid is difficult to use. In Yuviler-Gavish and Naseraldin's study [43], as mentioned before, a simple aid did not result in less reliance. It appears that people weigh their confidence in their own ability to handle the task against the time and effort they will have to invest to use the decision aid, and having previous experience with the task shifts the balance toward the former.

For the second study, the results demonstrated that in the second session, participants who were given the simple aid and did not get any aid before were willing to rely it. In contrast to our hypothesis, however, participants under the "hard first, simple second" contrast effect condition relied on the simple decision aid less than participants without this manipulation did. An additional analysis showed that the participants under this contrast effect condition continued to rely on the hard-to-use decision aid in the second session.

Adherence to the hard-to-use algorithm even when a simpler one is suggested might be explained by the "*sunk cost effect*." The sunk cost effect is the increased tendency to persist in an endeavor once an investment of money, effort or time has been made. The effect leads to irrational behavior since only marginal costs and benefits, and not past costs, should influence the decision making [52]. The sunk cost effect has been demonstrated in many fields. For example, Staw and Hoang [53] showed that the amount teams spent for players in the National Basketball Association (NBA) influenced how much playing time players got and how long they stayed with NBA franchises. McCarthy, Schoorman and Cooper [54] demonstrated the escalation of commitment in venture capital investment as a representative case of the sunk cost effect. This effect was also demonstrated in theater attendance: people who bought tickets to a show will not miss it even if they really do not want to come, e.g., in the situation of a storm [52].

According to the sunk cost effect, it may be that participants who had already invested time and effort in learning how to use the hard-to-use algorithm are less likely to give it up and use a simpler algorithm even if it is clear that the latter algorithm will save them time. Hence, benchmarking

a simple decision aid under previous experience with a harder to use one did not lead to the expected consequence of increasing reliance on the simple decision aid. Future research could focus on the contrast effect, but with an effort to reduce the sunk cost effect. For example, giving participants a less hard-to-use decision aid as the benchmark, which may reduce their investment and, in turn, the sunk cost effect, or emphasize that the simple algorithm can produce at least the same performance results.

As for task performance, in the first study adopting the algorithm did not result in significantly better performance. As we noted above, the algorithm was not designed to produce the same results as perfect knowledge of customers' future orders, and the difference between expected performance with perfect use of the algorithm and the best possible performance was quite large (mean total costs of 32.6 and 5.9 per day, respectively). Indeed, our findings point to the potential pitfalls of encouraging decision makers to use non-optimal decision aids and particularly the risk of overreliance [15, 16]. Interestingly, however, we found no evidence of overreliance on the algorithm in our study. On the contrary, the performance of the Aid group was better than the performance the group would have achieved with perfect use of the algorithm: 31.7 and 27.6 for the first and second sessions, respectively. In contrast, the No Aid group consistently scored worse than they would have with perfect use of the algorithm, at 35.3 and 38.1 for the first and second sessions, respectively.

We assume that participants who used the algorithm generalized their experience with it to develop an even better strategy, which improved their performance. More precisely, we postulate that while the algorithm used in the present study was not optimal, it had the effect of moderating variation in participants' orders, and participants in the Aid group may have learned to apply this strategy. It is even possible that given additional time, the improvement in their performance would have been statistically significant.

For the second study, task performance in the second session was similar for the Simple Aid and Hard-Simple Aid groups, and better than that of the No Aid group. In other words, both groups who got the aids were able to improve their performance with these aids, but the investment was different. Participants who were exposed to the hard-to-use decision aid in the first session continued to use it even after exposure to a much simpler aid. Participants who were exposed only to the simple decision aid adopted it. Hence, both groups used the aids and improved their performance, but the Hard-Simple Aid group did not take the chance to reduce its effort by moving to the simple decision aid.

The limitations of the current research are several. First, the number of participants in each study was limited, and hence, it is possible that the results were affected, and small changes in pattern could have been statistically significant if the number had been increased. Second, although trying to balance the percentage of males in each of the groups, since there might be gender differences in relying on a decision aid, the balance was not perfect. Third, the participants were all young engineering students from the same college, and it is possible that performing the study with different populations would have changed the results. And lastly, the external validity of the research is not very high, since the task was laboratory-based and only partially imitated the real-world conditions.

To conclude, in the current work we contributed to the scarce literature of manipulating the contrast effect in relying on a decision aid in DSS, by examining two experience-based contrast effects. We showed that these contrast effects can affect the reliance on a decision aid and task performance, but the effects might not be straightforward, and several other factors might contribute to the results produced. When a decision aid is hard to use, previous experience with the task will reduce reliance on such an aid once it is introduced. At the same time, we showed that at least in some cases, even when a decision aid is known not to be optimal, experience with such a decision aid can help users improve their own strategies. In addition, other effects such as sunk cost can cause users to adopt undesired behaviors, e.g., adherence to the harder decision aid instead of improving efficiency by relying on an easier one when it is offered. The current results should be examined further in future research using better decision aids, which may improve performance, and for other domains and tasks. In addition, since the task was laboratory-based and we did not tackle real-world problems, an enhancement to more realistic tasks is needed in the future.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflict of interest.

Availability of data and materials All data and material will be supplied upon request.

Consent to participate All participants gave a signed consent.

Consent for publication The authors have a consent for publication.

Ethics approval This research was approved by ORT Braude Ethics Committee.

References

1. Alvarado-Valencia JA, Barrero LH (2014) Reliance, trust and heuristics in judgmental forecasting. Comput Hum Behav 36:102–113
2. Blattberg RC, Hoch SJ (1990) Database models and managerial intuition: 50% model + 50% manager. Manag Sci 36(8):887–899
3. Hale DP, Kasper GM (1986) The effect of human–computer interchange protocol on decision performance. J Manag Inform Syst 6(1):5–20
4. Hoc J (2001) Towards a cognitive approach to human–machine cooperation in dynamic situations. Int J Hum Comput Stud 54(4):509–540
5. Saunders-Newton D, Scott H (2001) "But the computer said!": credible uses of computational modeling in public sector decision making. Sci Comput Rev 19(1):47–65
6. Zhang P, Li N (2004) An assessment of human–computer interaction research in management information systems: topics and methods. Comput Hum Behav 20(2):125–147
7. Creaser JI, Rakauskas ME, Ward NJ, Laberge JC, Donath M (2007) Concept evaluation of intersection decision support (IDS) system interfaces to support drivers' gap acceptance decisions at rural stop-controlled intersections. Transp Res Pt F-Traffic Psychol Behav 10(3):208–228
8. Preston H, Storm R, Donath M, Shankwitz C (2004) Review of Minnesota's rural crash data: methodology for identifying intersections for intersection decision support (IDS). Minnesota Department of Transportation, Minneapolis
9. Esmaeilzadeh P, Sambasivan M, Kumar N, Nezakati H (2015) Adoption of clinical decision support systems in a developing country: antecedents and outcomes of physician's threat to perceived professional autonomy. Int J Med Inf 84(8):548–560
10. O'Sullivan D, Fraccaro P, Carson E, Weller P (2014) Decision time for clinical decision support systems. Clin Med 14(4):338–341
11. Portela F, Aguiar J, Santos MF, Silva Á, Rua F (2013) Pervasive intelligent decision support system—technology acceptance in intensive care units. Advances in information systems and technologies. Springer, Berlin, pp 279–292
12. Romano MJ, Stafford RS (2011) Electronic health records and clinical decision support systems: impact on national ambulatory care quality. Arch Intern Med 171(10):897–903
13. Dulcic Z, Pavlic D, Silic I (2012) Evaluating the intended use of decision support system (DSS) by applying the technology acceptance model (TAM) in business organizations in Croatia. Proc Soc Behav Sci 58:1565–1575
14. Shim JP, Warkentin M, Courtney JF, Power DJ, Sharda R, Carlsson C (2002) Past, present, and future of decision support technology. Decis Support Syst 33:111–126
15. Layton C, Smith PJ, McCoy CE (1994) Design of cooperative problem-solving system for en-route flight planning: an empirical evaluation. Hum Factors 36:94–112
16. Parasuraman R, Molloy R, Singh IL (1993) Performance consequences of automation-induced complacency. Int J Aviat Psychol 3:1–23
17. Dzindolet MT, Pierce LG, Beck HP, Dawe LA (2002) The perceived utility of human and automated aids in a visual detection task. Hum Factors 44:79–94
18. Parasuraman R, Riley V (1997) Human and automation: use, misuse, disuse, abuse. Hum Factors 39:230–253
19. Riley V (1996) Operator reliance on automation: theory and data. In: Parasuraman R, Moulouas M (eds) Automation and human performance: theory and applications. Lawrence Erlbaum, Mahwah, pp 19–35
20. Lee JD, See KA (2004) Trust in automation: designing for appropriate reliance. Hum Factors 46:50–80
21. Dixon SR, Wickens CD, McCarley JS (2006) How do automation false alarms and misses affect operator compliance and reliance? In: Proceedings of the human factors and ergonomics society, 50th annual meeting, pp 25–29
22. Dixon SR, Wickens CD, McCarley JS (2007) On the independence of compliance and reliance: are automation false alarms worse than misses? Hum Factors 49:564–572
23. Dzindolet MT, Peterson SA, Pomranky RA, Pierce LG, Beck HP (2003) The role of trust in automation reliance. Int J Hum Comput Stud 58:697–718
24. Gao J, Lee JD (2006) Extending decision field theory to model operator's reliance on automation in supervisory control situations. IEEE Trans Syst Man Cybern Part A Syst Hum 36:943–959
25. Lee J, Moray N (1992) Trust, control strategies and allocation of function in human–machine systems. Ergonomics 35:1243–1270
26. Lee JD, Moray N (1994) Trust, self-confidence, and operators' adaptation to automation. Int J Hum-Comput Stud 40:153–184
27. Muir BM (1989) Operators' trust in and use of automatic controllers in a supervisory process control task. Doctoral thesis, University of Toronto
28. Spain RD, Madhavan P (2009) The role of automation etiquette and pedigree in trust and dependence. In: Proceedings of the human factors and ergonomics society annual meeting, vol 53, no. 4. Sage, pp 339–343
29. Lacson FC, Wiegmann DA, Madhavan P (2005) Effects of attribute and goal framing on automation reliance and compliance. In: Proceedings of the human factors and ergonomics society annual meeting, vol 49, no. 3. Sage, pp 482–486
30. Dijkstra JJ, Liebrand WB, Timminga E (1998) Persuasiveness of expert systems. Behav Inf Technol 17(3):155–163
31. Madhavan P (2014) Is ignorance bliss? Role of credibility information and system reliability on user trust in emergent technologies. Adv Cogn Eng Neuroergon 11:1532–1539
32. Herr PM, Sherman SJ, Fazio RH (1983) On the consequences of priming: assimilation and contrast effects. J Exp Soc Psychol 19(4):323–340
33. Thornton B, Maurice J (1997) Physique contrast effect: adverse impact of idealized body images for women. Sex Roles 37(5–6):433–439
34. Kobre KR, Lipsitt LP (1972) A negative contrast effect in newborns. J Exp Child Psychol 14(1):81–91
35. Chen C, He G (2016) The contrast effect in temporal and probabilistic discounting. Front Psychol 7:304
36. Dawes RM, Singer D, Lemons F (1972) An experimental analysis of the contrast effect and its implications for intergroup communication and the indirect assessment of attitude. J Pers Soc Psychol 21(3):281
37. Di Lollo V, Beez V (1966) Negative contrast effect as a function of magnitude of reward decrement. Psychon Sci 5(3):99–100
38. Oikawa H, Sugiura M, Sekiguchi A, Tsukiura T, Miyauchi CM, Hashimoto T, Kawashima R (2012) Self-face evaluation and self-esteem in young females: an fMRI study using contrast effect. Neuroimage 59(4):3668–3676
39. Schuh AJ (1978) Contrast effect in the interview. Bull Psychon Soc 11(3):195–196
40. Smither JW, Reilly RR, Buda R (1988) Effect of prior performance information on ratings of present performance: contrast versus assimilation revisited. J Appl Psychol 73(3):487
41. Rice S, Trafimow D, Clayton K, Hunt G (2008) Impact of the contrast effect on trust ratings and behavior with automated systems. Cogn Technol J 13(2):30–41
42. Yang XJ, Wickens CD, Hölttä-Otto K (2016) How users adjust trust in automation: contrast effect and hindsight bias. In: Proceedings of the human factors and ergonomics society annual meeting, vol 60, no. 1. Sage, Los Angeles, pp 196–200

43. Yuviler-Gavish N, Naseraldin H (2016) The effect of previous experience when introducing a decision aid in a decision support system for supply chain management. Cogn Technol Work 18(2):439–447

44. Fan X, Oh S, McNeese M, Yen J, Cuevas H, Strater L, Endsley MR (2008) The influence of agent reliability on trust in human–agent collaboration. In: Proceedings of the 15th European conference on cognitive ergonomics: the ergonomics of cool interaction. ACM, p 7

45. Hoff KA, Bashir M (2015) Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors 57(3):407–434

46. Sanchez J, Rogers WA, Fisk AD, Rovira E (2014) Understanding reliance on automation: effects of error type, error distribution, age and experience. Theor Issues Ergon Sci 15(2):134–160

47. Jackson GC, Taylor JC (1998) Administering the MIT Beer game: lessons learned. Dev Bus Simul Exp Learn 25:208–214

48. Liu H, Howley E, Duggan J (2009) Optimisation of the Beer Distribution Game with complex customer demand patterns. In: IEEE congress on evolutionary computation, CEC'09. IEEE, pp 2638–2645

49. Ravid G, Rafaeli S (2000) 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

50. Sterman J (1984) Instructions for running the Beer Distribution Game. Systems Dynamics Group, Sloan School of Management, Cambridge

51. XJ Technologies©. www.anylogic.com. The Beer Distribution Game. https://cloud.anylogic.com/model/b0156f6d-6c04-431b-b48d-1b875b2720e7?mode=SETTINGS

52. Arkes HR, Blumer C (1985) The psychology of sunk cost. Organ Behav Hum Decis Process 35:124–140

53. Staw BM, Hoang H (1995) Sunk costs in the NBA: why draft order affects playing time and survival in professional basketball. Adm Sci Q 40:474–494

54. McCarthy AM, Schoorman FD, Cooper AC (1993) Reinvestment decisions by entrepreneurs: rational decision-making or escalation of commitment? J Bus Ventur 8:9–24