

How Correct and Defect Decision Support Systems Influence Trust, Compliance, and Performance in Supply Chain and Quality Management

A Behavioral Study Using Business Simulation Games

Philipp Brauner^{1,2(✉)}, André Calero Valdez^{1,2}, Ralf Philipsen^{1,2}, and Martina Ziefle^{1,2}

¹ Human-Computer Interaction Center at RWTH Aachen University, Aachen, Germany

² Chair of Communication Science, RWTH Aachen University, Aachen, Germany
{brauner, calero-valdez, philipsen, ziefle}@comm.rwth-aachen.de

Abstract. Supply Chains and production networks are complex sociotechnical cyber-physical systems whose performance is determined by system, interface, and human factors. While the influence of system factors (e.g., variances in delivery times and amount, queuing strategies) is well understood, the influence of interface and human factors on supply chain performance is currently insufficiently explored. In this article, we analyze how performance is determined by the correctness of Decision Support Systems and specifically, how correct and defect systems influence subjective and objective performance, subjective and objective compliance with the system, as well as trust in the system. We present a behavioral study with 50 participants and a business simulation game with a market driven supply chain. Results show that performance (−21%), compliance (−35%), and trust (−25%) is shaped by the correctness of the system. However, this effect is only substantial in later stages of the game and occluded at the beginning. Also, people’s subjective evaluations and the objective measures from the simulation are in congruence. The article concludes with open research questions regarding trust and compliance in Decision Support Systems as well as actionable knowledge on how Decision Support Systems can mitigate supply chain disruptions.

Keywords: Compliance · Trust · Decision support system · Supply chain management · Enterprise resource planning · Human factors · Business simulation game · Sociotechnical Cyber-Physical systems · Internet of production

1 Introduction

Global sourcing, increased competition, shorter innovation cycles, increased customers’ demand on product variety and quality, and shorter ramp-up processes challenge the effectiveness of increasingly complex and globally dispersed cross-company supply chains [1–4]. Manufacturing companies therefore seek for methods to understand and manage their operation’s stability, performance, and overall resilience [5]. We consider supply chains (SC) as complex socio-technical cyber-physical systems whose resilience, performance, and stability

is determined by system factors (e.g., delivery times or economic stability), human factors (e.g., ability to cope with variances in processes, understanding of the underlying system), and interface factors (e.g., data presentation, Decision Support Systems). Considerable efforts have been invested in understanding and reducing the complexity that arises from the system factors, such as Lean Manufacturing, shortening the length of the SC, or the reduction of the SC's complexity [1, 2, 6, 7]. However, the influences from interface and human factors on supply chain performance is currently insufficiently explored and therefore neither adequately addressed in teaching and vocational training, in the strategic design of supply chains, and the design and evaluation of enterprise resource planning systems.

Therefore, the following article presents a behavioral experiment that investigates the influence of the interface, namely of correct and defect Decision Support System (DSS), on compliance with the system, trust, decision efficacy, and overall supply chain performance.

In the remainder of this article Sect. 2 presents the related work on Supply Chain Disruptions, Decision Support Systems, and Business Simulations and Business Simulation Games. Section 3 describes our research model and operationalizes the investigated variables. Section 4 then presents the results of our empirical study. Section 5 concludes that adequate Decision Support Systems can mitigate the effect of supply chain disruptions and can therefore strengthen the resilience of the production network. The final Sect. 6 outlines the limitations of the study and a future research agenda.

2 Related Work

This section of related work presents the causes of supply chain disruptions in Sect. 2.1, Decision Support Systems in Sect. 2.2, and Business Simulation games in Sect. 2.3.

2.1 Supply Chain Disruptions

Supply chain disruptions can be triggered by a variety of causes ranging from unexpected demand spikes, industrial accidents, strikes, terror attacks, wars, or natural disasters. A systematic review of causes for supply chain disruptions can be found in Snyder et al. [1]. A prominent example for a disruption is the *bullwhip effect* or *Forrester effect* [8]: A singular variance in the customer's order in combination with insufficient communication upstream the supply chain is amplified at each tier and yields in stock level graphs that look like a bullwhip. Although identified and formalized over 50 years ago, this effect is still frequently discussed [4, 9, 10].

Methods for mitigating supply chain disruptions are manifold, but most focus on organizational aspects of the production network. Examples are the *postponement strategy* that increases the sourcing potential by increasing compatibility with other suppliers, the term *strategic stocks* refers to the concept of additional safety stock inventories to compensate demand fluctuations, or changes to the *pricing strategy* and other methods to redirect demand to products less affected by disruptions [6].

Blackhurst et al. identified that research on supply chain disruptions provides many high level, but only limited practical information on preventing and handling disruptions. Using semi-structured interviews and focus groups they studied the source for supply chain disruptions and focused on the three areas of *disruption discovery*, *disruption recovery*, and *supply chain redesign* [5]. A key finding is the importance of visibility and predictive analysis of potential supply chain disruptions by operatives. Specifically, they state that human operators have limited abilities to process the enormous amount of information available today and are therefore limited in their ability to detect upcoming disruptions. They suggest an automated supply chain intelligence that triggers human intervention after certain thresholds have been reached. This relates to the idea of Decision Support Systems presented in the next section.

2.2 Decision Support Systems

Precursors of Decision Support Systems (DSS) have been developed since the 1950's and 1960's and they aim at harnessing the computational power and storage abilities of computers to automate the programmable part of operational, tactical, or strategic decision problems [11, 12]. This part is usually routine, repetitive, structured and therefore easily solved by computers. The systems encode knowledge, models, and decision rules in a computable form and provide support through querying systems, reports, or visualizations to decision makers. These can integrate the results into the non-programmable part of the decision problem, which is often new, creative, ill-structured, or difficult to solve. Data warehouses [12], OLAP [13], and data-mining [14] are modern forms of DSSs and artificial intelligence is gaining importance due to its ability to facilitate processing of a large amount of fuzzy information [15]. Summarizing, adequately designed DSS are a necessity to enable decision makers to handle the growing amount of information and complexity and to facilitate the success of the Industrial Internet and Industry 4.0 [16, 17].

Ben-Zvi used a business simulation game to engage students in the development and use of Decision Support Systems in an educational setting and investigated their perceived usefulness and their relationship to performance [18, 19]. The study found that perceived benefits of using a DSS, user satisfaction, and performance of the simulated company are strongly related. Also, support systems with higher complexity yielded in higher company performance. Although situated in an entrepreneurial context, neither the influence of supply chain effects, nor the influence of deliberately defect DSS's were investigated.

Brauner et al. investigated the influence of a correct and defective DSS compared to no DSS (baseline) in regard to decision efficiency (speed) and effectivity (correctness) in a table reading task of limited complexity [20]. As expected, a correct DSS increased speed and accuracy of the task compared to the baseline experiment. In contrast, a defective DSS had a devastating effect on task accuracy, whereas the speed was only mildly affected. Thus, the defective support system annihilates the subjects' task accuracy, despite knowing about its defectiveness. Strikingly, the devastating effect only emerged for more complex tasks, whereas it could be compensated in easier settings. This study is the basis for the experiment presented in this article.

2.3 Business Simulation Games

Simulated business and supply chains are an established method to identify and quantify supply chain disruptions, to convey knowledge and expertise about supply chain management and material disposition, as well as to study human decision making in controlled experimental, although sufficiently complex scenarios [21]. An early example are behavioral studies on the Beer Distribution Game by Sterman [22]. They found that a singular increase in customer's demand is amplified upstream the supply chain, yielding in the well-known bullwhip effect described above. Later, Lee et al. [4] identified the processing of demand signals, rationing of the inventory, order batching, as well as price fluctuations as the key causes for the emergence of the bullwhip effect. Furthermore, Wu and Katok investigate the influence of learning and communication on the bullwhip effect and found that experience alone is not sufficient for reducing the effect, but that collaboration and communication in combination with expertise reduces the order amplification [23]. Sarkar and Kumar investigated the effect of upstream (i.e., from the retailer) and downstream (i.e., from the supplier) disruptions and weather sharing knowledge about the disruption mitigates its effect in a behavioral experiment [24]. For upstream events (i.e., disruptions at the manufacturer) sharing information lead to a reduction in variances and overall supply chain costs, whereas limited effects were found for sharing information about downstream disruptions (e.g., at the retailer).

We developed the "*Quality Intelligence Game*", a sophisticated simulation model embedded in a turn-based business game rooted in Forrester's Beer Distribution Game [8] (focus on multi-echelon effects) and Goldratt's game (focus on quality variances along the supply chain) [25]. Players are part of market driven supply chain and must invest in the internal production quality, the incoming goods inspection, and the procurement of supplies. They must infer the current state of the production from the presented data and then need to find an optimal tradeoff between these three measures. Neglecting at least one of the measures yields in poor performance, as delivery bottlenecks or poor product quality are punished by the customer.

The underlying supply chain simulation model and the game's user interface can be experimentally controlled to investigate the influence of supply chain disruptions, unexpected changes to supplier's quality, or to the presentation of the company's various metrics, such as stock level, costs for quality inspection, and customers' complaints, or other KPIs. The average product quality or the attained performance can then serve as a benchmark for evaluating learning interventions, the influence of system complexity, or changes to the user interface.

3 Research Design

To understand the influence of the correctness of the decision support system on trust, compliance respectively use of the system, and overall performance, we applied a three-stage experimental design (Fig. 1 illustrates the research design):

User Factors:
Explanatory Variables

Interface factors:
Correctness of the DSS, randomized within-subject variable

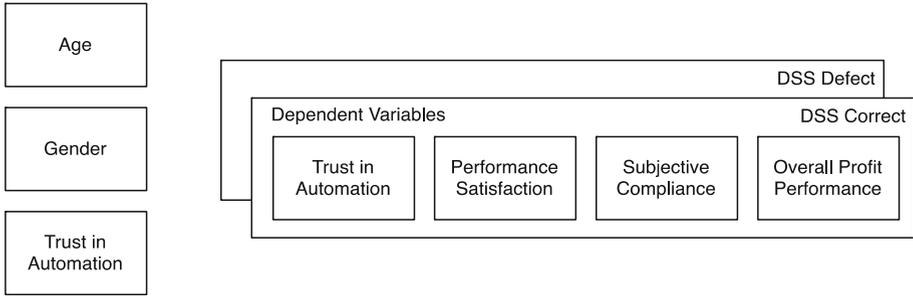


Fig. 1. Visualization of the experimental plan.

First, a pre-questionnaire captures the subjects' demographic data, trust in automation using a generic scenario (i.e., an app that suggests the number of beverages to buy for a party). Second, the participants played two rounds of the “Quality Intelligence Game” business simulation game—16 turns each—described above (without artificially induced supply chain disruptions) and log files capture company, interaction, and performance metrics. Third, a final questionnaire measures the subjects' evaluation of the perceived game performance, the trust towards the DSS, as well as the perceived compliance with the system for each of the two rounds. Unless otherwise noted, all subjective measures are captured on 6-point Likert scales from 0 to 5 (max.) and are rescaled to 0% to 100%.

Explanatory user factors: *Trust in Automation* (TiA) is captured on the scale by Jian et al. [26]. To measure the individuals' generic trust towards a support system we let them evaluate a fictitious app for planning the number of beverages to buy for a party. Despite the scenario based approach, the scale achieved an excellent internal reliability ($\alpha = .804$, 12 items).

Within-subject factors: The *Correctness* of the Decision Support System is the within-subject factor and the players randomly started either with a *defect* or a *correct* DSS in the first round of the game. In the *correct* condition, the DSS suggested very good, although not perfect orders for all turns of the game. In the *defect* condition, the suggestions of the DSS were about 50% below the value of the correct condition (easy to perceive as the suggestion is way below the customer's order and the penalties skyrocket in the following turns). To give the participants a false sense of security, the DSS always started with correct suggestions for the first five turns; then it switched to *defect* mode until the end of the game. In contrast to previous studies on this game, no disruptions were investigated as within-subject factors (cf., [27]).

Dependent variables: The trust in the DSS was captured after each round with the *Trust in Automation* scale from above [26]. The participants' compliance with the DSS is captured using a *subjective compliance* Likert scale ranging from 0% to 100%.

To understand the influence of the correctness of the DSS on performance, we captured subjective and objective performance measures. Based on Goldratt and Cox, the *company profit* is calculated as the cumulated net profit for each round of the game [25]. In addition, the subjects reported on their subjective *performance satisfaction* and the *subjective relative performance* compared to other players on a 6-point Likert scale ranging from “not satisfied” to “very satisfied”.

Methods: The results are analyzed with parametrical and non-parametrical methods, using bivariate correlations (Pearon’s r or Spearman’s ρ), Wilcoxon tests, single, and repeated multi- and univariate analyses of variance (M/ANOVA), and multiple linear regressions. The type I error rate (level of significance) is set to $\alpha = .05$ (findings . $05 < p < .1$ are reported as marginally significant). Pillai’s value is considered for the multivariate tests and effect sizes are reported as η^2 . If the assumption of sphericity is not met, Greenhouse-Geisser-corrected values are used, but uncorrected *dfs* are reported for legibility. As the performance from the simulation model is not normally distributed ($KS-Z_{round1} = 1.946$, $KS-Z_{round2} = 2.054$, $p < .001$) analyses of this model are performed with non-parametrical tests. Whiskers in diagrams represent the standard error (SE), arithmetic means are reported with standard deviations (denoted \pm).

3.1 Description of the Sample

40 people (23 male, 17 female) aged 20–56 ($M = 28.5 \pm 8.6$) years participated voluntarily in the web-based study and completed both rounds of the game (from 54 participants in the first round 25% did not completed second round). The initial Trust in Automation (TiA) had an average score of $73.0 \pm 13.9\%$ (0–100% max.) and it was neither related to age, nor gender.

4 Results

The results section is structured as follows: First, the link between objective measures from the simulation model and the participant’s subjective responses is established. Second, as the experimental setup taints the effect of practice and learnability with the effect of the correctness of the decision support system, the effect of correctness is discussed for each of the two consecutive rounds individually (between-subject). Third, a brief evaluation of the influence of practice and the communalities between both rounds is presented. Forth, the effect of correctness of the DSS is analyzed for both rounds combined (within-subject).

4.1 Preface: Congruency of System and Subjective Measures

The results show a strong relationship between the measures captured in the simulation game and the participant’s subjective responses. Thus, users are able to estimate how well they have performed.

Most importantly, the data shows a strong relationship between the performance measured in the simulation model and the performance satisfaction in the first ($\rho_{n=57} = .669, p < .001$) and second round of the game ($\rho_{n=44} = .300, p = .048 < .05$), as well as the perceived relative performance in the first ($\rho_{n=54} = .460, p < .001$) and second round ($\rho_{n=40} = .609, p < .001$). Hence, player's that reported a high performance and high satisfaction were actually good in the game.

Correspondingly, the number of order changes in the game's user interface is also strongly *negatively* related to the subjective compliance in the first ($\rho_{n=54} = -.721, p < .001$) and second round of the game ($\rho_{n=37} = -.755, p < .001$). Hence, participants who followed the suggestions of the DSS made less changes to the orders and reported a higher compliance with the system. On average, the number of order changes in the first round is 9.2 ± 6.1 and 9.8 ± 6.6 in the second round of the game compared to a maximum of 18 possible changes. The reported compliance is at $37.0 \pm 28.1\%$ respectively $45.7 \pm 30.5\%$ and thus in the same range as the measured compliance.

As subjective and objective measures behave similarly, the following sections focus on the subjective measures reported by the participants of the study. This facilitates the use of the more powerful parametrical methods for analyzing the study, despite the non-parametrical measurements from the simulation model.

4.2 Independent Evaluation of Both Rounds

This section illuminates the effect of the DSS' correctness independently for the first and second round of the game (i.e., neglecting influences of repetition).

In the first round, participants with a correct DSS achieved a higher *overall profit* ($Md = 11075$), higher *performance satisfaction* ($61.4 \pm 30.3\%$), and higher *relative performance* ($52.6 \pm 21.6\%$) than participants with a defective DSS ($Md = -29825, 51.0 \pm 33.6\%, 46.7 \pm 22.9\%$). Likewise, the reported and measured *compliance* with the system is higher for the correct system ($41.8 \pm 30.1\%, Md = 9.5$) compared to the defective system ($32.5 \pm 28.8\%, Md = 10.5$). The *Trust in Automation* score is also higher for the correct system ($66.2 \pm 15.9\%$) than for the defective system ($56.9 \pm 20.5\%$). However, despite all measures tending towards a positive effect of a correctly working DSS, the effect is merely significant for the *overall profit* (see Table 1 and Fig. 2, left).

Table 1. Effect of a Decision Support System's correctness in the first round of the game.

Overall profit	Order changes	Subjective compliance	Performance satisfaction	Relative performance	Trust in automation
Defect DSS:					
Md = -29825 M = -20708 ± 34644	Md = 10.5 M = 10.1 ± 5.5	32.5 ± 25.8%	51.0 ± 33.6%	46.7 ± 22.9%	59.9 ± 20.5%
Correct DSS:					
Md = 11075 M = 3274 ± 16627	Md = 9.5 M = 9.2 ± 6.7	41.8 ± 30.1%	61.4 ± 30.3%	52.6 ± 21.6%	66.2 ± 15.9%
U = 508.5 p = .027 < .05*	MW-U = 681.5 p = .673 > .05	F = 1.492 p = .227 > .05	F = 1.501 p = .226 > .05	F = .959 p = .332 > .05	F = 3.574 p = .064 > .05

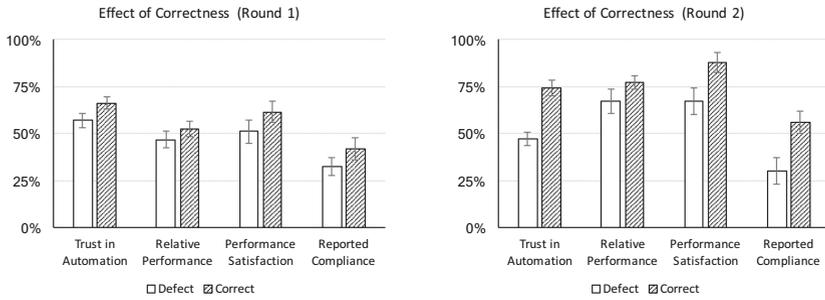


Fig. 2. Effect of a correct and defect DSS on *Trust in Automation*, *Performance Satisfaction*, *Subjective Relative Performance*, and *Subjective Compliance* for the first and second round of the game (whiskers indicate the SE).

In the second round, the effect of the correctness of the DSS are much clearer. Even though the difference is not significant, the *overall profit* is higher ($Md = 11675$) for the correct DSS than for the defective DSS ($Md = -10150$). Likewise, the measured and reported *compliance* with the correct DSS is higher ($56.0 \pm 28.6\%$, $Md = 7$) than for the defect DSS ($30.0 \pm 26.9\%$, $Md = 14$).

The correct DSS is also positively influencing the *performance satisfaction*, which is significantly higher for the correct system ($87.8 \pm 24.7\%$) than for the defect system ($67.3 \pm 33.5\%$). Although the difference is not significant, a similar—though smaller—effect seems to emerge for the subjective *relative performance* ($77.1 \pm 17.1\%$ vs. $67.0 \pm 28.5\%$). Consequently, the *Trust in Automation* is also significantly higher for the correct system ($74.3 \pm 20.7\%$) than for the defect system ($47.2 \pm 15.7\%$). Table 2 and Fig. 2, right shows these effects.

4.3 Effect of Repetition and Learnability

On average, the *overall performance* of the first and second round of the game were strongly related ($\rho = .751$, $p < .001$), but without a significant increase in attained performance ($Z = -.132$, $p = .895 > .05$). This suggests two conclusions: First, that some participants consistently play good, whereas others play bad. Second, that the

Table 2. Effect of a Decision Support System’s correctness in the second round of the game.

Overall profit	Order changes	Subjective compliance	Performance satisfaction	Relative performance	Trust in automation
Defect DSS:					
$Md = -10150$ $M = -12928 \pm 32007$	$Md = 14.0$ $M = 10.9 \pm 7.0$	$30.0 \pm 26.9\%$	$67.3 \pm 33.5\%$	$67.0 \pm 28.5\%$	$47.2 \pm 15.7\%$
Correct DSS:					
$Md = 11675$ $M = 6059 \pm 11284$	$Md = 7.0$ $M = 8.7 \pm 6.1$	$56.0 \pm 28.6\%$	$87.8 \pm 24.7\%$	$77.1 \pm 17.1\%$	$74.3 \pm 20.7\%$
$MW-U = 309.0$ $p = .056 > .05$	$MW-U = 364.5$ $p = .283 > .05$	$F = 7.807$ $p = .008 < .05^*$	$F = 5.516$ $p = .024 < .05^*$	$F = 1.934$ $p = .172 > .05$	$F = 23.494$ $p < .001^{**}$

influence of the DSS's correctness is rather strong and diminishes the influence of practice or learnability identified in earlier work [28].

Furthermore, the *order changes* in the first and second round of the game are positively related ($\rho_{n=49} = .557, p < .001$), which again indicates that some participants are more likely to adjust the order levels suggested by the DSS than others.

A RM-MANOVA with the *game* (round 1 and round 2) as within-subject factor and *Trust in Automation*, *Relative Performance*, *Performance Satisfaction*, and *Compliance* as dependent variables revealed an overall significant effect ($F_{2,29} = 12.267, V = .629, p < .001$). Neither *Trust* ($F_{1,32} = .077, p = .784 > .05$) nor the reported *Compliance* ($F_{1,32} = 1.007, p = .323 > .05$) are significantly different after the first and second round of the game. Yet, significant effects emerge for *relative performance* ($F_{1,32} = 39.871, p < .001$), as well as for the *performance satisfaction* ($F_{1,32} = 13.444, p = .001$). *Relative performance* increases from 49.6% to 72.2% and *performance satisfaction* increases from 56.1% to 77.8%. Figure 2 left illustrates the influence of repetition.

4.4 Influence of the Defect Decision Support System

In addition to the findings present in Sect. 4.2 this section now analyses the influence of the DSS with a focus on the within-subject factor correctness (neglecting a possible influence of practice).

On average, the attained *performance* with a correctly working DSS was higher (4479 ± 14523) than with a defect DSS (-17275 ± 33486) and this difference is significant ($Z = -2.647, p = .008 < .05$). Also, the number of order changes for the correct DSS is slightly lower (9.0 ± 6.4) than for the defect DSS (10.4 ± 6.1). Yet, this difference is only marginally significant ($Z = -1.893, p = .058 < .1$).

Based on the congruence of the objective measures from the simulation model and the subjective measures established in Sect. 4.1, the following sections investigate the subjective measures using parametrical methods.

A RM-MANOVA with *Correctness* as within-subject variable and *Trust in Automation*, *Compliance*, *Performance Satisfaction*, and *Subjective Relative Performance* as dependent variables revealed a strong and significant overall effect ($V = .471, F_{4,29} = 6.455, p < .001, \eta^2 = .471$). *Correctness* significantly influences all four considered dependent variables, namely *Trust in Automation* ($F_{1,32} = 21.670, p < .001, \eta^2 = .404$), *Compliance* ($F_{1,32} = 4.643, p = .039 < .05, \eta^2 = .127$), *Performance Satisfaction* ($F_{1,32} = 8.274, p = .007 < .05, \eta^2 = .205$) and *Subj. Relative Performance* ($F_{1,32} = 7.386, p = .011 < .05, \eta^2 = .188$).

Specifically, the reported *Trust* in the correct DSS ($69.8 \pm 2.6\%$) was sig. higher than the reported *Trust* in the defect system ($52.6 \pm 2.7\%$). Accordingly, the reported *compliance* was also higher for the correctly working system ($48.4 \pm 4.3\%$) compared to the lower *compliance* with the defective system ($31.6 \pm 4.0\%$). Likewise, a correct DSS yields in higher perceived *relative performance* ($63.3 \pm 3.3\%$) and higher *performance satisfaction* ($73.3 \pm 4.3\%$) compared to the defect system ($55.3 \pm 4.0\%$, resp. $58.0 \pm 4.8\%$). Figure 3 shows these significant effects.

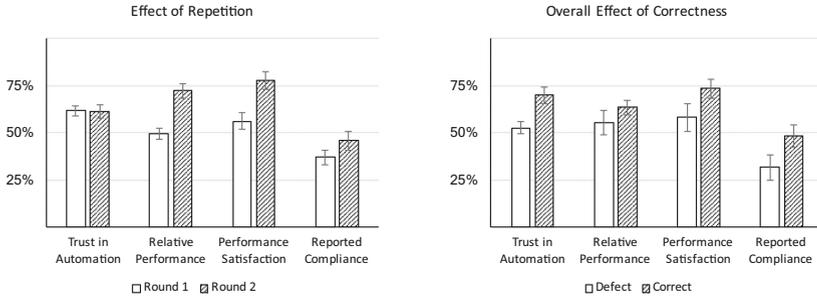


Fig. 3. Effect of practice (round 1 vs. round 2) and overall correctness (round 1 + 2) on *Trust in Automation*, *Relative Performance*, *Performance Satisfaction*, and *Subjective Compliance* (whiskers indicate the SE).

To understand if *Trust*, *performance*, *compliance*, and *profit overall* are interrelated and if this interrelationship is influenced by the correctness of the system, the following paragraphs present a correlation analysis of these four measures.

Correct DSS: For the correct Decision Support System, there are strong and significant relationships between the *Trust* in the system and the reported *compliance* ($\rho_{n=48} = .343, p = .017 < .05$), *relative performance* ($\rho_{n=47} = .550, p < .001$), and *performance satisfaction* ($\rho_{n=50} = .519, p < .001$). As expected, the relationship between subjective *relative performance* and *performance satisfaction* is also very high ($\rho_{n=48} = .716, p < .001$). However, the reported *compliance* is unrelated to *relative performance* ($\rho_{n=45} = .164, p = .281 > .05$) and *performance satisfaction* ($\rho_{n=48} = .042, p = .777 > .05$). Figure 4 (left) illustrates these relationships.

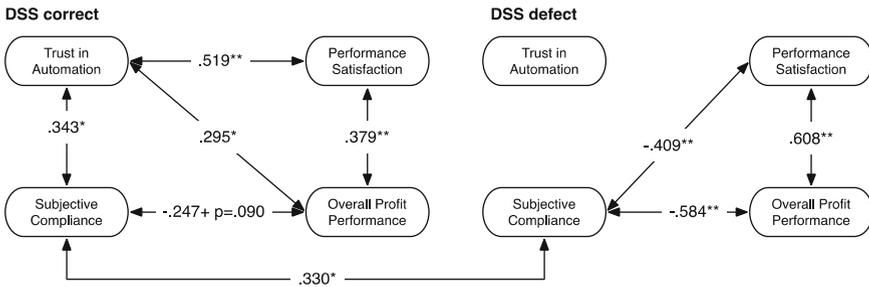


Fig. 4. Spearman’s ρ correlations between *Trust in Automation*, *Performance Satisfaction*, *Subjective Compliance* and attained *Overall Profit* for the *correct* and the *defect* DSS.

Defect DSS: For the defect Decision Support System, the reported *Trust* is neither related to the reported *compliance* ($\rho_{n=41} = .177, p = .268 > .05$), the *relative performance* ($\rho_{n=44} = -.170, p = .271 > .05$), nor the *performance satisfaction* ($\rho_{n=48} = -.109, p = .462 > .05$). The reported *compliance* is negatively associated with *relative performance* ($\rho_{n=40} = -.317, p = .047 < .05$) and the *performance satisfaction* ($\rho_{n=43} = -.409,$

$p = .006 < .05$). Again, subjective *relative performance* is strongly related with *performance satisfaction* ($\rho_{n=47} = .777, p < .002$). Figure 4 (right) presents the interrelationships for the defect Decision Support System.

Surprisingly, the generic *Trust in Automation* is neither related to the *Trust* in the correct system ($r = .173, p = .246 > .05$), nor to the *Trust* in the defect system ($r = .089, p = .567 > .05$), nor are the *Trust* in the correct and the defect system related ($r = .254, p = .100 > .05$). Also, neither subjective *relative performance* ($\rho_{n=40} = .193, p = .233 > .05$), nor *performance satisfaction* ($\rho_{n=44} = .016, p = .918 > .05$) are associated across both rounds. However, the reported *compliances* for both DSSs (defect, correct) are positively related ($\rho_{n=46} = .330, p = .046 < .05$).

5 Discussion

Our study provides some valuable insights regarding the positive influence of correctly working Decision Support Systems on performance, compliance, and trust, the harmful effects of defective DSS, as well as some methodological tidbits that may guide future research on business simulation games, Decision Support Systems, and human-factors in complex sociotechnical cyber-physical systems.

5.1 Benefits of Decision Support Systems

The study show that correctly working Decision Support System have an apparent positive influence on trust in a support system, compliance with the system, thus also on overall perceived and actual performance. Compared to the defective systems, the participants reported higher trust levels, a higher compliance, as well as higher performance satisfaction and most importantly, they also realized higher cumulated company profits. In summary, correctly working support systems are a valuable tool to relieve workers from repetitive or difficult tasks and increase their overall efficiency, as well as the overall efficiency of the manufacturing company.

5.2 Risks of Decision Support Systems

While the finding that a correct Decision Support System yields in a higher company profit seems trivial, the reverse perspective deserves attention: Although the subjects of the presented study must have noticed the defect of the DSS (the suggested orders were *clearly* below the customer's demand and the penalties increased), they still have followed the suggestion of the system to some extent, which has diminished the overall profit of the company, as well as the subjective performance.

This finding relates well to a study published earlier that illuminated the influence of corrected and defective DSSs in less complex table reading tasks [20]. However, the previous study concluded that the negative influence of defective DSSs on effectivity emerges only for more complex tasks, as defectiveness can easily be compensated for simple tasks. In contrast, this study investigated the influence of correctness in context and in a complex environment, but without an experimental consideration of task

complexity. Consequently, future work must address how correctness and defectiveness of Decision Support Systems influences efficiency, effectivity, and trust in relationship to complexity of the simulated environment.

In summary, defective support systems have an overall negative effect on work's efficiency, and thus a negative effect on the overall performance of companies and cross-company supply chains as complex sociotechnical cyber-physical systems.

5.3 Correct vs. Defect Decision Support Systems

For the case of a correct DSS, the study identified a higher trust in the automated system, as well as a positive relationship between trust and the compliance with the system, the satisfaction with the attained performance, as well with the actual performance. On the contrary, if the DSS is defect, trust is significantly lower and trust is independent to the compliance, satisfaction with the attained performance, as well as the actual company profit.

Surprisingly, there is a moderate negative association between the compliance with the system and the performance satisfaction in the case of a defect support system. Meaning that the participants complying with the system have noticed the defectiveness and their poor performance (hence the lower performance satisfaction and lower overall profit). Still, it is unclear why they followed the system's orders and under which conditions and when they would have started to neglect the system's suggestions. Interestingly, there was no relationship between compliance and performance satisfaction for the case of a correct Decision Support System. We conclude that people complying with a correct system do not feel the same level of accomplishment as people and may attribute their performance to the support system and not to their own individual abilities. Future work should therefore more closely address the role of attribution and Attribution Theory (cf. [29]) in regard to compliance with Decision Support Systems, performance satisfaction, as well as attained performance.

Interestingly, the reported compliance levels for the defect and the correct DSS are moderately related. This indicates that some subjects are more inclined to comply with the system and obey its orders than others. This raises the questions which and why operators in cyber-physical production systems are more likely to abdicate orders of the DSS than others, how operators can be trained to detect and disobey defective systems, and how trust in the system can be reestablished after such an incident.

A remarkable trifle of the study is the negative association of compliance with the system and the attained overall profit for both, the correct (although not significant) and the defect system. We assume that this relationship is caused by people focusing solely on the decisions support system and neglected other parts of the business simulation game. For the case of the defective system, compliance with the faulty suggestions obviously has a devastating effect. For the case of the correct system, focusing solely on the support system and thereby neglecting other parts of the simulated company also yields in lower profits, as managing the order levels was just one of three tasks in the game (for an isolated perspective of a single task see [20]).

5.4 Methodological Contributions

From the methodological perspective, the study revealed that objective measures from the simulation model are in accordance with the subjective measures reported by the participants, that the subjective measures require calibration through training, and that we identified a possible lower barrier for the applied trust scale.

Methodologically, the study established a strong relationship between the various perceived measures of the study (e.g., performance and compliance) and the objective measures captured in the underlying simulation of business game. As most objective efficiency and effectivity measures from simulations or actual production environments do not follow parametrical distributions, their statistical analysis—especially if combined with elements from psychometrical measures as in this study—is impeded. However, due to this strong relationship of objective and subjective measures, future studies can build on the analyses of the parametrical subjective measures which extends the methodological portfolio by more sophisticated statistical methods.

Apparently, a correctly working Decision Support System has a profound positive influence on objective and subjective *company profit*, objective and subjective *compliance*, as well as in *trust* in the system. Still, these effects are only discernible, yet not significant, for the first round of the game. Only in the second round of the game, these effects gain in power and yield in statistically significant results. We assume, that this is caused by a missing internal calibration and anchoring of the respective measures for first round of the game. In the subsequent round, a reference frame is established and yields in more separated measures, lower spreads, and clearer results. Therefore, future research addressing Trust, compliance, or perceived performance must ensure that an adequate reference frame or anchoring is established by providing training sessions or repeated measurements.

An additional trifle is a step towards the calibration of psychometric trust scales. Due to our empirical methodology with the induced malfunction of the DSS, we have measured the lower barrier of the trust scale by Jian et al. [26]. Due to the defectiveness of the DSS, it is rather unlikely that avg. trust scores will fall below the value of 52.6% on this scale. However, our approach is unsuitable to identify an upper barrier of this scale, as the current rating is not only affected by the correctness of the DSS, but probably also tainted by the effect of the underlying simulation model and the individual's abilities to cope with the supply chain's complexity. Future work will therefore have to address methods to empirically determine the scale's upper barrier.

In contrast to the compliance with the system and despite its high internal reliability, the three measurements of *Trust in Automation* were not related. This indicates that Trust in Automation—as least as captured in this study—is not an individual personal trait, but rather a state that is heavily influenced by the automated system and the reliability of this automation.

5.5 Summary

The study has shown that correctly working Decision Support Systems do have a positive influence on the effectivity of the simulated cross-company supply chain. We therefore

conclude that adequately designed Decision Support Systems can mitigate supply chain disruptions and that DSSs are one fundamental pillar to strengthen the resilience of manufacturing companies and likely other complex sociotechnical cyber-physical systems. Yet, defective Decision Support System can have a devastating effect on the performance, which was also identified for different settings in previous research [20]. Consequently, future research must identify how supply chain operators and workers in material disposition can be trained to notice defective Decision Support Systems, disobey their orders, and act successfully despite the lack of decision support.

Concluding, by empowering operators to harness the benefits of correctly working decision support and to mitigate the drawbacks of defective support systems, the overall performance of cross-company supply chains can be increased, their resilience can be strengthened, and their overall viability be established.

6 Outlook

In the current experiment the effect of the correctness of the Decision Support System is not clearly separated from and thereby tainted by the effect of repeating several rounds of the game (e.g., practice, fatigue, motivational change) discovered in earlier studies (e.g., [27]). Consequently, a follow up study with a significantly larger sample size should investigate this effect and must separate, identify, and quantify the influence of these factors.

Furthermore, the negative influence of a defective DSS has been shown in an abstract experimental setting and in this study with the business simulation game. Future studies must therefore investigate the influence of defective support systems in more complex or realistic settings. Hereto, the business simulation game's adjustable complexity might be used to help to increase the understanding of the interaction of interface and system complexity.

Finally, strategies and trainings must be identified, developed, and evaluated that enable operators to recognize defective or defective support systems and therefore prevent blind obedience of these systems.

Acknowledgements. We thank all participants for their willingness to contribute to our research and our colleagues Sebastian Stiller, Marco Fuhrmann, Hao Ngo, Robert Schmitt, for support and in-depth discussions on this work. Furthermore, we like to thank Sabrina Schulte for her research support. The German Research Foundation (DFG) founded this project within the Cluster of Excellence „Integrative Production Technology for High-Wage Countries” (EXC 128) and the integrated cluster domain ICD-D1 [30].

References

1. Snyder, L.V., Atan, Z., Peng, P., Rong, Y., Schmitt, A.J., Sinoysal, B.: OR/MS models for supply chain disruptions: a review. *IIE Trans.* **48**, 89–109 (2016)
2. Trent, R.J., Monczka, R.M.: Pursuing competitive advantage through integrated global sourcing. *Acad. Manage. Executive* **16**, 66–80 (2002)

3. Brauner, P., Philipsen, R., Fels, A., Fuhrmann, M., Ngo, H., Stiller, S., Schmitt, R., Ziefle, M.: A game-based approach to meet the challenges of decision processes in ramp-up management. *Qual. Manage. J.* **23**, 55–69 (2016)
4. Lee, H.L., Padmanabhan, V., Whang, S.: Information distortion in a supply chain: the bullwhip effect. *Manage. Sci.* **43**, 546–558 (1997)
5. Blackhurst, J., Craighead, C.W., Elkins, D., Handfield, R.B.: An empirically derived agenda of critical research issues for managing supply-chain disruptions. *Int. J. Prod. Res.* **43**, 4067–4081 (2005)
6. Tang, C.: Robust strategies for mitigating supply chain disruptions. *Int. J. Logistics* **9**, 33–45 (2006)
7. Blum, M., Runge, S., Groten, M., Stiller, S.: Interrelationships between product quality and different demand cases in ramp-up scenarios. *Procedia CIRP* **20**, 81–84 (2014)
8. Forrester, J.W.: *Industrial Dynamics*. MIT Press, Cambridge (1961)
9. Sarkar, S., Kumar, S.: Demonstrating the effect of supply chain disruptions through an online beer distribution game *. *Decis. Sci. J. Innovative Educ.* **14**, 25–35 (2016)
10. Brauner, P., Runge, S., Groten, M., Schuh, G., Ziefle, M.: Human factors in supply chain management – decision making in complex logistic scenarios. In: Yamamoto, S. (ed.) HIMI 2013. LNCS, vol. 8018, pp. 423–432. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-39226-9_46](https://doi.org/10.1007/978-3-642-39226-9_46)
11. Gorry, G.A., Morton, M.S.S.: A framework for management information systems. *Sloan Manage. Rev.* **13**, 50–70 (1971)
12. Kimball, R., Ross, M.: *The data warehouse toolkit: the complete guide to dimensional modelling*. Wiley, New York (1996)
13. Codd, E., Codd, S., Salley, C.: *Providing OLAP to User-Analysts: An IT Mandate* (1993)
14. Bra, A., Lungu, I.: Improving decision support systems with data mining techniques. In: *Advances in Data Mining Knowledge Discovery and Applications*. InTech (2012)
15. Phillips-Wren, G.: AI tools in decision making support systems: a review. *Int. J. Artif. Intell. Tools* **21** (2012)
16. Brauner, P., Ziefle, M.: Human factors in production systems – motives, methods and beyond. In: Brecher, C. (ed.) *Advances in Production Technology*. LNPE, pp. 187–199. Springer, Cham (2015). doi:[10.1007/978-3-319-12304-2_14](https://doi.org/10.1007/978-3-319-12304-2_14)
17. Calero Valdez, A., Brauner, P., Schaar, A.K., Holzinger, A., Ziefle, M.: Reducing complexity with simplicity - usability methods for industry 4.0. In: 19th Triennial Congress of the International Ergonomics Association (IEA 2015), Melbourne, Australia (2015)
18. Ben-Zvi, T.: The efficacy of business simulation games in creating Decision Support Systems: an experimental investigation. *Decis. Support Syst.* **49**, 61–69 (2010)
19. Ben-Zvi, T.: Measuring the perceived effectiveness of decision support systems and their impact on performance. *Decis. Support Syst.* **54**, 248–256 (2012)
20. Brauner, P., Calero Valdez, A., Philipsen, R., Ziefle, M.: Defective still deflective – how correctness of decision support systems influences user’s performance in production environments. In: Nah, F.F.-H., Tan, C.-H. (eds.) *HCIBGO 2016*. LNCS, vol. 9752, pp. 16–27. Springer, Cham (2016). doi:[10.1007/978-3-319-39399-5_2](https://doi.org/10.1007/978-3-319-39399-5_2)
21. Brauner, P., Ziefle, M.: How to train employees, identify task-relevant human factors, and improve software systems with business simulation games. In: Dimitrov, D., Oosthuizen, T. (eds.) *Proceedings of the 6th International Conference on Competitive Manufacturing 2016 (COMA 2016)*, pp. 541–546. CIRP, Stellenbosch (2016)
22. Sterman, J.D.: Modeling managerial behavior: misperceptions of feedback in a dynamic decision making experiment. *Manage. Sci.* **35**, 321–339 (1989)
23. Wu, D.Y., Katok, E.: Learning, communication, and the bullwhip effect. *J. Oper. Manage.* **24**, 839–850 (2006)

24. Sarkar, S., Kumar, S.: A behavioral experiment on inventory management with supply chain disruption. *Int. J. Prod. Econ.* **169**, 169–178 (2015)
25. Goldratt, E.M., Cox, J.: *The goal: a process of ongoing improvement*. North River Press, Great Barrington (1992)
26. Jian, J.-Y., Bisantz, A.M., Drury, C.G.: Foundations for an empirically determined scale of trust in automated system. *Int. J. Cogn. Ergon.* **4**, 53–71 (2000)
27. Philipsen, R., Brauner, P., Stiller, S., Ziefle, M., Schmitt, R.: Understanding and supporting decision makers in quality management of production networks. In: *Advances in the Ergonomics in Manufacturing. Managing the Enterprise of the Future 2014: Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics, AHFE 2014*, pp. 94–105. CRC Press, Boca Raton (2014)
28. Stiller, S., Falk, B., Philipsen, R., Brauner, P., Schmitt, R., Ziefle, M.: A game-based approach to understand human factors in supply chains and quality management. *Procedia CIRP* **20**, 67–73 (2014)
29. Niels, A., Guczka, S.R., Janneck, M.: The impact of causal attribution s on system evaluation in usability tests. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3115–3125 (2016)
30. Schlick, C., Stich, V., Schmitt, R., Schuh, G., Ziefle, M., Brecher, C., Blum, M., Mertens, A., Faber, M., Kuz, S., Petruck, H., Fuhrmann, M., Luckert, M., Brambring, F., Reuter, C., Hering, N., Groten, M., Korall, S., Pause, D., Brauner, P., Herfs, W., Odenbusch, M., Wein, S., Stiller, S., Berthold, M.: Cognition-enhanced, self-optimizing production networks. In: Brecher, C., Özdemir, D. (eds.) *Integrative Production Technology - Theory and Applications*, pp. 645–743. Springer International Publishing, Heidelberg (2017)