Effects of Visualization Techniques on Understanding Inconsistencies in Automated Decision-Making

Carl Corea
Institute for IS Research
University of Koblenz-Landau
ccorea@uni-koblenz.de

Sabine Nagel
Institute for IS Research
University of Koblenz-Landau
snagel@uni-koblenz.de

Patrick Delfmann
Institute for IS Research
University of Koblenz-Landau
delfmann@uni-koblenz.de

Abstract

The automation of business processes and decision-making has received major interest from practice and academia. As automation allows us to execute more processes (cases), monitoring automated decision-making is currently evolving into a big data analytics problem for companies. Thus, not only monitoring insights themselves, but also an effective use of such insights become important. In this context, the speed and ability to interpret data is closely related to the visualization of metrics and data. While various approaches for quantitative insights on automated decision-making have been proposed, there is currently no evidence as to how the specific visualization of such metrics helps companies to create more value from their data. In this report, we therefore present the results of an empirical experiment analyzing the cognitive effects of different visualization techniques for quantitative insights on understanding inconsistencies in automated decision-making data.

1. Introduction

Business process automation is a central challenge for today's businesses [1, 25]. In regard to the current digital transformation, company activities have to be digitalized and automated in order to stay competitive [11]. Thus, major efforts are currently being directed towards executing business processes via workflow management systems (WFMS) [30]. Such systems can automatedly perform predefined business processes, e.g., retail or customer service processes, to increase efficiency and reduce the manual effort needed in company activities. A central challenge here is automated decision-making, to allow WFMS to handle processes autonomously.

However, monitoring correct and consistent decision-making in WFMS is currently evolving into a big data analytics problem for companies [6, 25].

This becomes apparent in the scope of the *7V model* [13, 19], which indicates the main attributes of big data. For example, the *volume* of data created by WFMS is increasing rapidly. The online retailer Zalando, which handles customer processes with a WFMS [25], reports that 31 million cases were executed by their system only in the first quarter of 2019, which is nearly a 50% increase since the first quarter of 2017[1]. Also, as company processes span across systems and organizations, process data often includes heterogeneous and unstructured data, such as scanned documents. Thus, data *variety* and *veracity* also become increasingly challenging for companies [7, 23, 25]. Last, the *velocity* of data creation increases through automation. WFMS track processes in real-time, leading to shorter periods in which companies have to monitor decision-making.

In order to ensure consistent decision-making, companies must understand and utilize process case data to detect errors in automated decisions [6]. Such an understanding can be an important driver in creating value through innovation, e.g., by minimizing mistakes, improving WFMS, and streamlining business processes [6, 12, 19]. To support companies in this aim, the scientific field of inconsistency measurement has evolved and proposed so-called inconsistency measures, which are *metrics* that can help to identify those cases where inconsistencies have occurred in the decision-making [6, 8]. Also, inconsistency measures provide quantitative insights that can help to assess the severity of inconsistency and thus prioritize cases for the analysis by experts [6, 12, 24].

Due to the unique big data challenges arising in the context of WFMS, not only these quantitative metrics themselves, but also an effective *use* becomes even more important [19, 27]. In this context, the *speed and ability to interpret data* is closely related to a dimension of the 7V model, namely the *visualization* of metrics and data.

---

[1] https://zln.do/31wcTqz

While various approaches for inconsistency metrics have been proposed, there is currently no evidence as to how the specific *visualization* of such insights helps companies to create more *value* from their data. In this work, we thus investigate how visualization techniques for inconsistency metrics affect the capability to analyze and interpret inconsistencies in automated decision-making. To this aim, we introduce different visualization techniques for inconsistency metrics in section 2 and hypothesize the relation of these approaches and understanding inconsistencies in section 3. To verify our hypotheses, we conducted an empirical experiment using neurophysiological measurement, presented in sections 4-5. Our study suggests that providing users with a separate, ranked overview of metrics (cf. section 2) is associated with better understanding efficiency and less mental effort required to handle cases compared to other visualization techniques.

2. Background and Related Work

WFMS are systems that allow companies to integrate process- and decision logic, subsequently allowing them to execute business processes (semi-) automatically. If a new process (i.e., a case) is started, the WFMS conducts all tasks as predefined in the process model sequentially. During this traversal of the process model, decision-making is performed using business rules, which govern how the process should be executed based on the case data. Figure 1 shows an exemplary business process in the Business Process Model and Notation (BPMN)[2] and corresponding business rules in the Decision Model and Notation (DMN) standard[3].
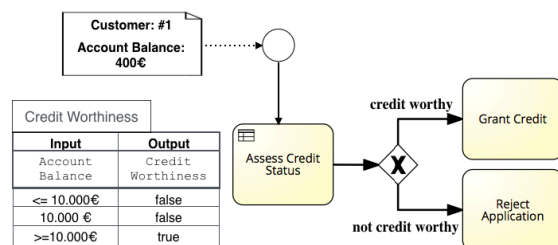

Figure 1. Exemplary business process and business rules

DMN allows formalizing rules with so-called decision tables. The rules in the shown table can be read such that "*if*" the account balance is <= 10.000,

---

"*then*" the customer is not credit worthy, and so forth. In Figure 1, given the shown case input, we thus see that the system can automatedly reject the request, reducing manual effort in decision-making.

However, business rules are created by human modelers, mostly collaboratively and incrementally [16, 17]. As suggested by a wealth of recent research, modeling errors can frequently occur in this setting [2, 3, 6, 24, 26]. For example, in a recent case study with a large insurance company, the authors in [2] found that 27% of rules had modeling errors.

In turn, such modeling errors can result in inconsistent decision-making [6]. For example, considering the business rules in Figure 1, the customer input of 10.000€ would result in contradictory conclusions due to erroneous modeling. Here, all rule conditions overlap, yielding contradicting conclusions. This flawed decision-making could thus result in compliance breaches, as the WFMS could perform unallowed activities. It is therefore essential for companies to monitor consistent decision-making during process execution as a driver for innovative and sustainable development of WFMS, e.g., by re-modeling business rules and improving operations [6, 10, 24].

Implementing such compliance monitoring is, however, becoming more and more difficult for companies due to the increasing amounts of data which are generated by WFMS [25]. Such data amounts are challenging in the scope of gaining business intelligence insights, as well as creating value from data and insights. For example, during compliance monitoring, where a case can consist of heterogeneous or even unstructured data, experts must analyze the entire case to correctly understand and resolve modeling errors.

To support companies in this aim, inconsistency metrics have been proposed, providing insights on inconsistencies in decision-making during run-time. Examples include [6] for an overview and [8, 29] for some recent surveys. An important family of such metrics are *culpability measures*, which aim to quantify the severity of inconsistency for *individual* business rules with a *numerical value*. The intuition is that a higher value reflects a higher severity of inconsistency. An exemplary culpability measure is the $C_\#$ measure [6], which assesses the culpability of an individual rule r by counting the number of other rules that contradict the respective rule r.

Example 1. *Revisiting the business rules from Figure 1 and a case input of 10.000€, this would mean that rule 3 would have a C#-value of 2, as the conclusion of rule 3 conflicts the two other rules. Respectively, rules 1 and 2 would each have a C#-value of 1, as they both individually contradict rule 3.*

Thus, culpability measures can pinpoint the actual causes of inconsistencies in WFMS cases as a driver for resolving modeling errors [6]. Nagel et al. (2019) could show that such inconsistency metrics are associated with a better understanding of inconsistencies. In light of the unique big data challenges in WFMS, not only these (quantitative) insights themselves but also the *speed of their use* become important to foster the innovative and sustainable development of WFMS [19]. Thus, the actual visualization of metrics becomes a critical success factor in creating value from case data, as it can impact how information can be analyzed and interpreted.

So far, there are two different visualization approaches for inconsistency metrics that have been proposed, namely an *integrated* visualization and a *ranked overview* visualization [6].

In the integrated approach, the culpability values are displayed directly *within* the respective decision table, i.e., next to the individual business rules. Figure 2 shows an example of such an *integrated* visualization for Example 1.

Figure 2. Integrated visualization

On the other hand, Corea et al. (2018) propose a separate *ranked overview* of inconsistency metrics. Instead of integrating the respective culpability values directly in the table, a reference (e.g., a colored dot) is used to point to a *separate*, *ranked* overview, where all rules are presented in a sorted list. Figure 3 shows such a ranked visualization for Example 1. Due to the ranking, it can be directly seen that rule 3 is the most problematic one and should be attended to first. Intuitively, such a ranked overview can provide benefits for larger tables, or if problems are distributed across different tables.

Figure 3. Ranking-based visualization

Regarding which of these visualization techniques is "better" to display big data insights, both approaches have advantages and disadvantages, and there are several contradicting aspects from the field of cognitive psychology that must be considered. An important aspect here is *cognitive load theory*, which describes the relation between cognitive load during information processing and the performance of understanding information [31].

In the *integrated* approach, the metric information and the rules are visually presented in a unified manner. This can lower the extraneous cognitive load for processing the information, due to a minimization of the so-called split-attention effect [31, 32, 33]. This effect can occur when two related pieces of information are visually distributed and need to be linked mentally by the expert, which is not the case in the integrated approach.

However, while this would advocate an integrated visualization, this approach can lead to a higher amount of case information which needs to be processed. It is important to realize that business rules are usually stored in multiple tables in practice. Thus, all tables need to be checked, and it is not sufficient to "simply" sort rules by their culpability values in the respective tables in an integrated visualization.

On the contrary, a *ranked overview* allows the expert to quickly comprehend which rules (and which decision tables) should be attended to first. Such a *recommendation* can thus guide modelers in handling cases and lower the amount of information that has to be processed. Also, results from cognitive psychology suggest that factors such as coherence or elaborative encoding of information can promote more efficient processing of information [20, 28].

Yet again, while this would advocate using a ranked visualization, the resulting split of information can be expected to introduce more cognitive load during information assimilation [31, 33]. As both proposed approaches have advantages and disadvantages, a further investigation based on literature is difficult at this point. Currently, there exists no *empirical* evidence investigating which of these visualization techniques helps companies to create more value from their data.

3. Research Aim

Following authors such as Surbakti et al. (2019) or Olszak & Zurada (2019), the effectiveness and speed of using insights into big data is a key factor in creating value for companies. As these abilities are closely linked to how information can be processed, this work aims at investigating how different visuali-

zation techniques for quantitative insights affect the effectiveness of their use and thus, also help companies to create more value from their data.

Accordingly, we derive the following research question:

RQ: How do visualization techniques for inconsistency metrics affect the capability to analyze and interpret inconsistencies in automated decision-making?

As introduced in the previous section, two visualization approaches have been proposed so far, namely an *integrated* approach and a *ranking-based* approach. Here, the ranking-based approach has been introduced more recently and makes use of culpability measures to present users with a recommendation. As a result, especially in real-life cases where experts would have to scan multiple and large tables, such a recommendation could potentially increase *efficiency* by guiding modelers. Accordingly, our first aim is to investigate how a ranking-based visualization affects understanding efficiency, i.e., the time needed for modelers to understand inconsistencies based on inconsistency metrics.

Hypothesis 1: A ranking-based visualization of inconsistency metrics is associated with a better understanding efficiency compared to an integrated visualization.

Also, while an integrated approach could potentially lower cognitive load due to a reduction of the split-attention effect, experts might struggle to gain a more holistic oversight of problems. Here, a ranked overview could potentially lower the mental effort needed to understand inconsistencies, as the ranking can be used to present a holistic prioritization of problems. Accordingly, our second aim is to investigate how a ranking-based visualization affects the objective mental effort needed for understanding inconsistencies.

Hypothesis 2a: A ranking-based visualization of inconsistency metrics is associated with less objective mental effort compared to an integrated visualization.

Next to the objective mental effort, the visualization might also affect the perceived ease of use, which could be used as a driver for the development of visualization techniques. Therefore, our third aim is to investigate the effects of visualization on perceived ease of use, i.e., the perceived mental effort needed to understand inconsistencies.

Hypothesis 2b: A ranking-based visualization of inconsistency metrics is associated with less perceived mental effort compared to an integrated visualization.

Despite the potential advantages of the ranking-based visualization, it is currently not clear whether the advantages of a ranked overview outweigh the potential cognitive costs that can be expected due to the split-attention effect [33]. Thus, empirical evidence is needed. Consequently, we opted for an experimental research approach to test these hypothesized relations.

We consequently follow the experimental research methodology as proposed by Neuman [18], as it is highly suitable for the investigation of causal relations between independent variables and their effects. Here, independent variables are manipulated in a controlled environment in order to assess the effects that follow the manipulation. We thus see this research methodology as highly appropriate, as the visualization techniques can be seen as the independent variables which the researcher can manipulate, and the methodology thus allows us to assess the effects of these techniques. The dimensions of understanding efficiency and mental effort are based on the experiment design in [32].

4. Experiment

We conducted an empirical experiment[4] to test our hypothesis. In the following, we describe our experiment design and the measures used to verify our hypotheses.

4.1. Experiment Design

In order to empirically evaluate the effects of the two approaches of displaying insights into WFMS data, we confronted participants with cases and corresponding questions covering inconsistencies.

We designed our experiment as a single factor experiment, as this allows us to assess the effects of a single factor on a shared variable [22]. In our case, the considered factor is the visualization approach of the inconsistency values, with the factor levels being "integrated-" and "ranking-based" visualization.

The experiment was split into two runs (i.e., models), which each represent a separate domain. We divided our participants into two groups randomly and tested them for both domains. Here, each domain was tested with different factor configurations for the

---

[4] The experiment can be downloaded from https://bit.ly/2KfsBAU

two groups, one seeing the inconsistency metrics directly in the table and the other one having them visualized in a separate ranking. As a result of this, the experiment is balanced with repeated measurement. This has the advantage of the participants using both factor levels without being exposed to the same cases twice. This repeated measurement thus allows each participant to generate more data, which can lead to more precise, powerful results [4] and, therefore, increases generalizability. As the order of factor levels was reversed between groups, we could also counterbalance a potential distortion following [32].

Figure 4 gives an overview of the overall experiment design. In the first run, which contains tasks 1-4, Group 1 was exposed to the inconsistency metrics being shown directly in the tables (integrated visualization), while group 2 had them visualized in a separate ranking (ranking-based visualization). This situation was inverted in the second run, where Group 1 was provided with the inconsistency metrics in a separate ranking, while they were displayed directly in the tables for Group 2.
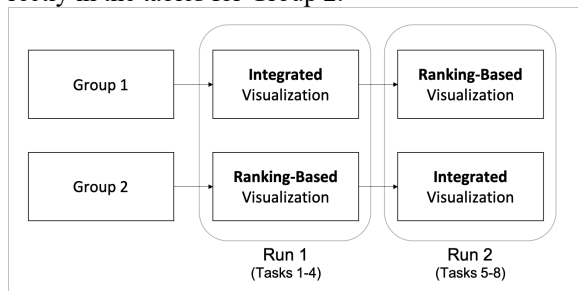

Figure 4. Experiment design

The entire experiment and the questions were formulated in English to warrant comprehension in the scope of reproducible research.

4.2. Experiment Structure & Instrumentation

Before exposing the participants to actual cases and corresponding questions, we presented them with an introduction to the experiment. These introductory slides covered all topics that were needed to be able to perform the tasks, such as the basics of decision management and DMN tables, as well as the quantitative measures and their two visualization approaches used in the experiment. The tutorial also covered an exemplary task to ensure that the participants became familiar with the quantitative measures and their use. The participants were able to go through the introductory slides at their own speed, as an understanding of the used concepts was crucial for the success of the experiment.

As shown in Figure 4, each run comprised four different tasks, each containing a case and a corresponding question. We designed the cases to cover basic types of inconsistencies in WFMS cases; however, there were no dependencies between cases.

In general, each task was divided into three, resp. four, areas as illustrated in Figure 5, which shows the structure of an exemplary task for both factor levels.
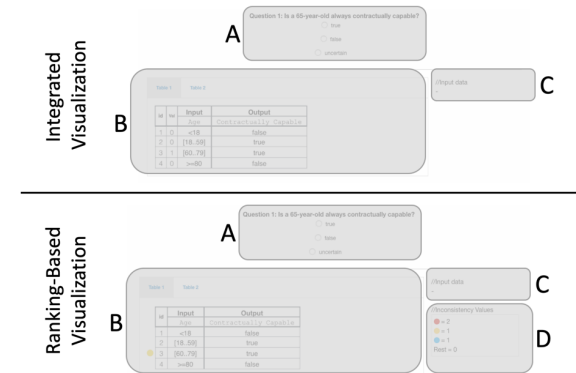

Figure 5. Exemplary task structure

The question and the corresponding response options are displayed at the top of the screen (A). Below, the current case is shown, which is divided into two areas. On the left, the first DMN table is shown (B). If the case consists of multiple tables, the user can switch between them using the tabs displayed above the table. The different tables in one case may be directly related, e.g., inconsistencies can be distributed among multiple tables. However, there is also the possibility of a table being irrelevant to the current question. Furthermore, the participant was shown case data, which is needed to answer the questions. The case data can either be implied by the question or listed in the corresponding box (C). Depending on the current factor level, the inconsistency values are either displayed to the right of the row ID, or in a separate, ordered ranking (D).

There were four different types of questions:
- Questions that asked about the existence of inconsistencies in the current case (e.g., "Is there an inconsistency in James' case?")
- Content-related questions (e.g., "Is the person in this case contractually capable?"). The questions could either be answered with "true", "false", or – if no conclusion could be made due to an existing inconsistency – "uncertain".
- Questions that asked for the rule with the highest/lowest number of inconsistencies (e.g., "Which rule is in conflict with the highest number of other rules for this case?").

- Questions that asked for a number of rules that contradict a specific number of other rules (e.g., "How many rules are in conflict with EXACTLY three other rules in this case?")

4.3. Measurements

To test our hypotheses, we used three different types of measurements.

Understanding efficiency was measured using the time from the point a case was displayed until the participant clicked on an answer. Next, we used the eye-fixation duration to measure objective mental effort. The eye-fixation duration is defined as the period of time where the eyes are focused on a specific location and therefore remain still. As vision is suppressed as soon as the eyes are moved, new information can only be captured during fixation [21]. Thus, fixation duration is an indication for the time it takes a participant to process the respective information. In turn, a higher fixation duration indicates a higher objective measure for mental effort [14]. In addition to the objective measures, we also asked the participants which run, i.e., which visualization approach they found easier to work with, in order to measure the perceived mental effort. We also conducted semi-structured interviews with all participants after the experiment, to gain further insights into the reasons behind their choices and their general impressions.

4.4. Settings

We used the cloud-based eye tracking software EYEVIDO lab[5] to create our study. While the introductory slides were added as pdf files, we implemented the tasks as HTML files. We also verified the clear visibility of all task components from a distance of at least 60cm during a pre-test (c.f. a further description of the pre-test in section 4.5.)

As already depicted in section 4.2, the screen was divided into up to four parts: the question area at the top of the screen and the corresponding case as well as the inconsistency metrics below. As most cases consisted of more than one table, the participants were able to click on the corresponding tab to switch between tables when needed. Furthermore, some tables required scrolling as their size prevented them from being displayed in full size.

We used a *myGaze n 30Hz* eye tracker in combination with a 22-inch screen and a resolution of 1680 x 1050. The experiment was conducted in an IT lab at the University of Koblenz-Landau. To keep the lighting situation consistent, the blinds were closed, and the ceiling light was the only source of light.

4.5. Participants

In order to ensure the understandability and readability of the introductory slides and experiment questions, we conducted a pre-test with 8 Ph.D. students prior to the actual experiment. During the pre-test, we found that one major problem was the lack of trust in the inconsistency metrics, as many participants preferred to double-check the implication of the quantitative insights manually. As a result of this, we refined the introductory slides in order to build more trust in the values, as our focus was an investigation of different visualization approaches, and the general advantage of the values themselves has already been evaluated in [15].

A total of *48* undergraduate and graduate students from the school of Computer Science at the University of Koblenz-Landau participated in our experiment. Please observe that this school does not only include traditional computer science but also includes institutes focusing on business informatics or business administration. Thus, the participants represented diverse fields. Although the introductory covered all relevant concepts, all participants had a general knowledge of business process management based on their study programs.

The assignment of the 48 students into the two groups was performed at random. Furthermore, no incentive was offered, so participation was voluntary.

5. Results

To test our hypotheses, we assessed the experiment data with the measures described in section 4.3. Given that our experiment was set up as a between-subject design experiment, we statistically compared the respective measures of the individual groups. Here, we proceeded as follows.

First, we checked if each dependent variable could be assumed to be normally distributed, using the Shapiro-Wilcox test at a significance level of 0.05. Then, if the data could be assumed to be normally distributed, we verified whether the dependent variables had an equal variance, using Levene's test[6] at a significance level of 0.05. Given a normal distribution and equal variance, we ran an independent-sample t-test. If data could not be assumed to be

---

normally distributed, we used the Wilcoxon-Mann-Whitney test[7]. For all comparison tests, we assumed the commonly used significance level of 0.05.

For Hypothesis 1, the measured time needed to answer the questions was normally distributed and met the assumption of equal variance for both models. We consequently ran an independent-sample t-test between groups 1 and 2 for both models, using the time needed to answer all four questions in one run as the dependent variable.

Table 1. Test of Hypothesis 1
(understanding efficiency)

	Run 1		Run 2	
Group	1 (*i*ntegrat-ed)	2 (*r*anked)	1 (*r*)	2 (*i*)
N	24	24	24	24
Mean	34.54	29.04	23.09	25.02
SD	8.71	12.89	8.46	16.24
p (1-tailed**)**	**0.048**		0.30	

As can be seen in Table 1, the group using a ranked overview had a significantly better understanding efficiency in run 1, which partially supports our hypotheses. For run 2, the difference was not significant, yet the group using the ranked visualization was still better on average.

Conclusion 1: For the measured data, a ranking-based visualization of inconsistency metrics is (partially) associated with a better understanding efficiency compared to an integrated visualization.

For Hypothesis 2a, the eye-fixation duration measured during the experiments was also normally distributed and met the assumption of equal variance for both models. Accordingly, we ran an independent sample t-test, using the eye-fixation duration needed for answering all four questions in one run as the dependent variable. For 6 participants, eye-tracking data could not be tracked due to an unknown system failure, thus N = 42 for comparing fixation duration.

Table 2. Test of Hypothesis 2a
(objective mental effort)

	Run 1		Run 2	
Group	1 (*i*)	2 (*r*)	1 (*r*)	2 (*i*)
N	21	21	21	21
Mean	30.95	22.86	19.58	19.78
SD	16.22	23.73	15.22	26.97
p (1-tailed**)**	**0.004**		0.47	

As depicted in Table 2, the group using a ranked overview had a significantly lower eye-fixation duration in model 1, which partially supports Hypothesis 2a. For model 2, the difference was again not significant, yet the eye-fixation duration was still lower on average for the group using a ranked overview.

Conclusion 2: For the measured data, a ranking-based visualization of inconsistency metrics is (partially) associated with less objective mental effort compared to an integrated visualization.

Table 3 shows the results of the perceived mental effort, as provided by the participants.

Table 3. Perception of mental effort

	Group 1	Group 2	Total
Ranking is easier	14	11	25 (52%)
Integration is easier	5	9	14 (29%)
Equal	5	4	9 (18%)

To recall, both groups performed two runs and were exposed to both visualization techniques. Overall, the ranking-based overview was perceived as easier by the majority of participants. To statistically compare the perceived mental effort, we coded the answers by participants as follows: If the participant selected that the ranking based visualization was easier, the ranking based visualization was assigned 2 points. Vice versa, if the participant selected that the integrated visualization was easier, the integrated visualization was assigned 2 points. If a participant indicated that both visualization techniques were perceived to require equal mental effort, both the ranking-based and integrated visualization were assigned 1 point. We then compared the average perceived mental effort between the two visualization forms. As the coded data is ordinal (i.e., possible answers were 0, 1 or 2), the data could not be assumed to be normally distributed. Since for the perceived mental effort *every* participant provided his or her answer based on *both* models, we ran the Wilcoxon signed-rank test[8] for the differences in the perceived mental effort for both visualizations.

Table 4. Test of Hypothesis 2b
(perceived mental effort)

	N	Coded mean	SD	P (1-tailed)
Ranking	48	1.22	0.88	**0.039**
Integrated	48	0.77	0.88	

---

[7] The Wilcoxon-Mann-Whitney test is used to compare the distribution data from two samples

[8] The Wilcoxon signed-ranked test is used for dependent samples

The perceived mental effort for the ranking-based visualization was significantly lower than for the integrated visualization, which supports our hypothesis.

Conclusion 3: For the measured data, a ranking-based visualization of inconsistency metrics is associated with less perceived mental effort compared to an integrated visualization.

6. Discussion

For run 1, our results fully support Hypotheses 1 and 2a, which suggests that the ranked overview visualization was associated with a better understanding efficiency and less objective mental effort needed for handling cases in this run. In run 2, the difference in group performances was not large enough to be statistically significant. Still, the group which used the ranking-based visualization had on average a better understanding efficiency and less mental effort than the group using the integrated visualization in run 2. We assume that the performances could have been affected by participants handling similar cases in run 1 and run 2, i.e., a learning effect could have affected the measured performance in run 2. However, the similar structure of run 1 and run 2 was necessary to mitigate a bias based on different questions. Intuitively, this would not be the case in practice, where experts face far more complex and individual cases. Thus, based on the statistical significance in run 1 and the fact that participants using a ranking-based visualization in run 2 were still on average better, we cannot reject our hypotheses. Also, our results fully support Hypothesis 2b (which assessed both runs), indicating that perceived mental effort was significantly lower for the ranking-based visualization, compared to an integrated visualization of metrics.

Figure 6 shows the distribution of average *time* needed to complete a run, as well as the corresponding average *fixation duration*.
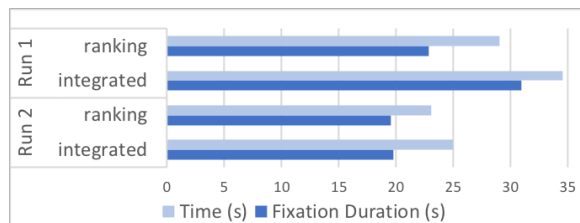

Figure 6. Distribution of time and fixation duration (units in seconds)

As can be seen, participants using a ranking-based visualization took less time and had a lower fixation duration in both runs. Especially in run 1, the proportion of fixation duration relative to total time needed

was also much higher for the integrated visualization (89% vs. 79%). Thus, the participants spent a higher fraction of time trying to understand information for the integrated visualization as opposed to the ranking-based visualization.

As mentioned, a split of information when using a separate ranking can increase the extraneous cognitive load during information assimilation [31, 32, 33]. Yet, in an integrated approach, more information has to be processed due to a lack of guidance. This can be visualized using the collected eye tracking data. Figure 7 shows a heatmap of the participants' gaze distribution. The colored areas indicate areas of visual focus, where the red color ("heat") represents a higher fixation duration. As can be seen, participants tried to process a higher amount of information in the integrated approach. Thus, the benefits of having to process less information seem to outweigh the cognitive costs of information assimilation (split-attention effect).
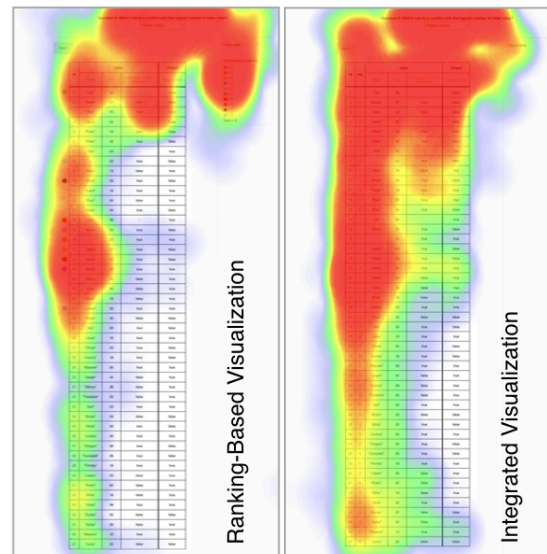

Figure 7. Heatmaps of gaze distribution for different visualization techniques

In the post-experiment interviews with the participants, we were also able to gain further insights into the reasons for the perceived mental effort and their general impression regarding the different visualization approaches. Most participants stated that they felt relieved that the ranking allowed them to find important information more efficiently. For example, participants described that they could focus on specific rows more effectively and were therefore often able to skip irrelevant parts or even entire tables when having access to the ranking.

Interestingly, some participants stated that despite being sure about understanding the provided metric information, they did not feel confident in using it right away. This resulted in the participants trying to re-verify the metric assessment manually. However, the participants stated an increase of trust during the course of the experiment. Although the general trust in the inconsistency metrics seems to have improved compared to the pre-test (due to the refinement of the introductory slides), trust in the insights seems to be an important issue when creating value from big data.

7. Conclusion

In this work, we investigated the relation of visualization techniques on understanding inconsistencies in automated decision-making. To this aim, we conducted and evaluated an empirical experiment using neurophysiological measurement. Our results indicate that a ranking-based visualization was (partially) associated with a better understanding efficiency and lower mental effort in handling inconsistent cases.

Our empirical insights contribute to the development of future WFMS and tools. Recommender systems and elaborate rankings based on quantitative metrics should be developed, as the benefits of having to process less information seem to outweigh the cognitive costs of information assimilation (split-attention effect). Also, this study affirms that the specific visualization of big data insights can help companies to use these insights more effectively. This capability thus represents an added value of big data and business intelligence in regard to innovation, i.e., improving WFMS or streamlining processes.

In the scope of the 7V model, the visualization component is often overlooked. More research is needed that investigates how the visualization of big data insights can help to foster innovation and help companies to create value from their data.

Due to the research methodology, our study is not without limitations. Intuitively, the results are dependent on the participant composition (e.g., number of participants or personality traits) which could affect external validity. Our results are based on experiment data from *48* participants, which we see as comparable to other related studies in the field, e.g. [22] (*19* participants), [14] (*23* participants), [9] (*28* participants), [32] (*50* participants) or [5] (*75* participants). Also, English was not the native language of the participants, which could affect internal validity. However, all participants indicated a sufficient understanding of the English language as a precondition for participating in the experiment. A further limitation is that the participant group comprised students, as opposed to domain experts, which

could also affect external validity. However, the study curricula include highly practice-based and related subjects, such as business process management or business administration. Also, as the cases included basic inconsistencies, a transferability to other domains seems plausible. In future work, we aim to conduct experiments with professionals and larger sample sizes.

Bearing these limitations in mind, we could, however, observe that participants using a ranking-based visualization of data insights could on average handle cases in a lower amount of time and with less mental effort needed. Therefore, our findings can help companies to create more value from their data by using these insights on visualization effects to guide the future development of WFMS and monitoring tools.

Future work should investigate more elaborate rankings and recommendation systems. For example, participants indicated that switching between the ranking and information in the tables, such as the row number, was perceived as difficult. Here, future work should investigate how to integrate more information into the ranking and exploit amenities of the integrated approach to develop advanced rankings or hybrid approaches. Also, some participants suggested providing additional guidance through further visual elements in the ranking, e.g., more elaborate or graphical encodings.

Our results affirm that the visualization of big data insights is a key factor in creating value from data for companies. The effects of visualization techniques should more strongly be included as a driver for the development of information systems. In this way, specific visualization techniques can be used to foster economic value by means of helping companies to understand their data, thus enabling innovation through improved decision-making and more sustainable WFMS management.

8. Acknowledgements

9. References

[1] Aalst, W.M.P. van der, M.L. Rosa, and F.M. Santoro, "Business Process Management", *Business & Information Systems Engineering 58*(1), 2016, pp. 1–6.

[2] Batoulis, K., and A. Nesterenko, "Decision Management in the Insurance Industry: Standards and Tools", *BPM*

2017 Industry Track co-located with the 15th International Conference on Business Process Management, Springer International Publishing (2017), 11.

[3] Calvanese, D., M. Dumas, Ü. Laurson, F.M. Maggi, M. Montali, and I. Teinemaa, "Semantics, analysis and simplification of DMN decision tables", *Information Systems 78*, 2018, pp. 112–125.

[4] Charness, G., U. Gneezy, and M.A. Kuhn, "Experimental methods: Between-subject and within-subject design", *Journal of Economic Behavior & Organization 81*(1), 2012, pp. 1–8.

[5] Chen, T., W. Wang, M. Indulska, and S. Sadiq, "Business Process and Rule Integration Approaches-An Empirical Analysis", *International Conference on Business Process Management*, Springer (2018), 37–52.

[6] Corea, C., and P. Delfmann, "Supporting Business Rule Management with Inconsistency Analysis", *Proceedings of the BPM*, (2018), 09–14.

[7] Goes, P.B., "Big data and IS research", *Mis Quarterly 38*(3), 2014, pp. iii–viii.

[8] Grant, J., and M.V. Martinez, *Measuring Inconsistency in Information*, College Publications, 2018.

[9] Haji, F.A., D. Rojas, R. Childs, S. de Ribaupierre, and A. Dubrowski, "Measuring cognitive load: performance, mental effort and simulation task complexity", *Medical Education 49*(8), 2015, pp. 815–827.

[10] Hashmi, M., G. Governatori, H.-P. Lam, and M.T. Wynn, "Are we done with business process compliance: state of the art and challenges ahead", *Knowledge and Information Systems*, 2018, pp. 1–55.

[11] Imgrund, F., M. Fischer, C. Janiesch, and A. Winkelmann, "Approaching digitalization with business process management", *Proceedings of the MKWI*, 2018, pp. 1725–1736.

[12] Lu, R., S. Sadiq, and G. Governatori, "Measurement of Compliance Distance in Business Processes", *Information Systems Management 25*(4), 2008, pp. 344–355.

[13] Manyika, J., M. Chui, B. Brown, et al., *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, 2011.

[14] Meghanathan, R.N., C. van Leeuwen, and A.R. Nikolaev, "Fixation duration surpasses pupil size as a measure of memory load in free viewing", *Frontiers in Human Neuroscience 8*, 2015.

[15] Nagel, S., C. Corea, and P. Delfmann, "Effects of Quantitative Measures on Understanding Inconsistencies in Business Rules", *Proceedings of the 52nd Hawaii International Conference on System Sciences*, (2019).

[16] Nelson, M.L., R.L. Rariden, and R. Sen, "A lifecycle approach towards business rules management", *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, IEEE (2008), 113–113.

[17] Nelson, M.L., and R. Sen, "Business rules management in healthcare: A lifecycle approach", *Decision Support Systems 57*, 2014, pp. 387–394.

[18] Neuman, W.L., *Basics of social research: Qualitative and quantitative approaches*, Pearson Education, Inc., 2007.

[19] Olszak, C., and J. Zurada, "Big Data-driven Value Creation for Organizations", *Proceedings of the 52nd Hawaii International Conference on System Sciences*, (2019).

[20] Park, B., T. Flowerday, and R. Brünken, "Cognitive and affective effects of seductive details in multimedia learning", *Computers in Human Behavior 44*, 2015, pp. 267–278.

[21] Rayner, K., "Eye Movements in Reading and Information Processing: 20 Years of Research", *Psychological Bulletin 124*(3), 1998, pp. 372–422.

[22] Reijers, H.A., J. Mendling, and R.M. Dijkman, "Human and automatic modularizations of process models to enhance their comprehension", *Information Systems 36*(5), 2011, pp. 881–897.

[23] Sackmann, S., S. Kuehnel, and T. Seyffarth, "Using Business Process Compliance Approaches for Compliance Management with Regard to Digitization: Evidence from a Systematic Literature Review", *International Conference on Business Process Management*, Springer (2018), 409–425.

[24] Sadiq, S., and G. Governatori, "Managing Regulatory Compliance in Business Processes", In *Handbook on Business Process Management*. 2015, 23.

[25] Schrepfer, M., M. Kunze, G. Obst, and J. Siegeris, "Why are process variants important in process monitoring? The case of Zalando SE", In *Business Process Management Cases*. Springer, 2018, 431–448.

[26] Smit, K., M. Zoet, and M. Berkhout, "Verification capabilities for business rules management in the Dutch governmental context", *2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*, IEEE (2017), 1–6.

[27] Surbakti, F.P.S., W. Wang, M. Indulska, and S. Sadiq, "Factors influencing effective use of big data: A research framework", *Information & Management*, 2019.

[28] Thalmann, M., A.S. Souza, and K. Oberauer, "How does chunking help working memory?", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2018.

[29] Thimm, M., "On the Compliance of Rationality Postulates for Inconsistency Measures: A More or Less Complete Picture", *KI - Künstliche Intelligenz 31*(1), 2017, pp. 31–39.

[30] Van Der Aalst, W., K.M. Van Hee, and K. van Hee, *Workflow management: models, methods, and systems*, MIT press, 2004.

[31] Wang, W., *Integrating Business Process Models and Rules*, Springer, 2019.

[32] Wang, W., M. Indulska, S. Sadiq, and B. Weber, "Effect of Linked Rules on Business Process Model Understanding", *International Conference on Business Process Management*, Springer International Publishing (2017), 200–215.

[33] Wang, W., M. Indulska, and S.W. Sadiq, "Cognitive Efforts in Using Integrated Models of Business Processes and Rules.", *CAiSE Forum*, (2016), 33–40.