

# Advancing Trust Visualisations for Wider Applicability and User Acceptance

Oksana Kulyk\*, Benjamin Reinheimer\*, Paul Gerber\*, Florian Volk<sup>†</sup>, Melanie Volkamer\*<sup>‡</sup>, Max Mühlhäuser<sup>†</sup>

\*<sup>†</sup>Technische Universität Darmstadt, Germany

<sup>‡</sup>Karlstad University, Sweden

\*name.surname@secuso.org; <sup>†</sup>{volk,max}@tk.tu-darmstadt.de

**Abstract**—There are only a few visualisations targeting the communication of trust statements. Even though there are some advanced and scientifically founded visualisations—like, for example, the opinion triangle, the human trust interface, and T-Viz—the stars interface known from e-commerce platforms is by far the most common one. In this paper, we propose two trust visualisations based on T-Viz, which was recently proposed and successfully evaluated in large user studies. Despite being the most promising proposal, its design is not primarily based on findings from human-computer interaction or cognitive psychology. Our visualisations aim to integrate such findings and to potentially improve decision making in terms of correctness and efficiency. A large user study reveals that our proposed visualisations outperform T-Viz in these factors.

**Index Terms**—Trust visualisation, user study, human-computer interaction

## I. INTRODUCTION

The research in computational trust has seen considerable advances in the last years and is becoming a known field of research, even outside its own community. Especially, computational trust has gained some popularity in the field of IT security; the gap between hard security [1] and applied security—that involves humans, which are inherently unmeasurable and oftentimes act beyond predefined scenarios—has been filled with mechanisms and insights from computational trust.

In the past, computational trust research focused on evolving and perfecting the model(s) of trust and on a deeper understanding of what trust is in its essence. While the modelling was conducted by researchers with a background in math and economics, the interpretation of trust has been conducted from the viewpoint of philosophy and sociology. Surprisingly, the interest by human-computer interaction communities has been low. As such, only a few examples of trust visualisations exist, e.g., [2]–[5].

### A. Trust Visualisations in the Wild

Most of the time, developers revert to simple visualisations, like the stars interface (a distinct number of stars alongside the amount of evidence this number is based on) known from e-commerce websites, to communicate trust statements. Common for this approach is a loss of semantics when communicating trust. By design, a trust statement is much more than an average of ratings. When using stars, aspects like the reliability of a statement are hidden from the human user. Furthermore, a stars

interface reduces the dimensionality of a rating, i.e., stars only communicate a single aggregated value. However, it would be possible to display a distinct stars interface for every dimension to overcome this drawback. This approach, however, is far from an integrated representation of rating dimensions.

The question why most real-world applications of trust continue to rely on simple visual cues like stars is one that the trust research community still has to answer.

At the 2012 Trust Management Symposium, Glass and Farmer addressed the question why the industry does not seem to be interested in advanced trust visualisations. The simple mechanisms described above (i.e., communication trust via stars) seem to be good enough for their specific applications. There is no interest in supplying human participants with more than this minimal information, which—coincidentally—is easy to comprehend even without any explanation on the internals of computational trust. We consider this answer to be correct, but nevertheless insufficient.

The stars interface and similar ones completely neglect the aspect of reliability of a communicated value. In popular trust models, this reliability is given by (un)certainty measures or the specific configuration of fuzziness parameters, e.g., by Jøsang in [6], by Ries in [7], or by Kant and Dwiwedi in [8]. Instead of fact-based proof of reliability, trust visualisations used in industry implicitly claim the correctness of their trust values by relying on what we want to call “authentic recommenders”, i.e., they claim that their trust sources are real, uninfluenced customers or independent experts. It is often impossible for end-users to verify such a claim.

### B. Further Advancing Trust Visualisations

It seems that trust visualisations that are developed from the viewpoint of a trust researcher and that carry insights into trust are still too complex for many applications that involve humans. Especially in e-business cases where trust is used to compare alternatives, trust communication must be easy to grasp and fast to compare as well as quick to understand. However, a visual trust statement still needs to confer a trust value and a reliability score instead of just a rating.

The aim of this paper is to continue the evolution of trust visualisations by improving the T-Viz visualisation, a multicriterial trust visualisation that incorporates a trust value and a certainty statement [5], [9]. T-Viz itself is based on earlier trust visualizations, mostly on the *Human Trust Interface* by

Ries et al. [3]. However, the design of T-Viz was not primarily driven by human-computer interaction design practices, but by the aim to display multiple criteria of a trust statement at once. Based on the psychological research in human-computer interaction and cognitive psychology, we designed two trust visualisations, which are based on the ideas of T-Viz, but aim at being easier and faster to comprehend by human users. Our hypotheses are investigated with the help of a user study with 101 participants. The evaluation we conduct is driven by established practices in psychological studies and supported by human-computer interaction design practices.

Our study reveals that the new proposed visualisations enable human users to better identify a higher and/or more reliable rating in comparison with T-Viz<sup>1</sup>. Moreover, human users are quicker when using the proposed visualisations.

### C. Structure

The structure of this paper is as follows: the following Section II reviews and categorizes scientific work related to ours. In Sections III and IV, we describe the three trust visualisations we compare, namely the original T-Viz, a version enhanced with star symbols, and one version enhanced with bar charts. Section V described the study design, while Sections VI and VII present and discuss the results we obtained. Finally, Section VIII concludes.

## II. RELATED WORK

Up until today, the most established visualisations are used in e-commerce scenarios. Such visualisations use an amount of stars or similar symbols to describe quality. These simple visualisations usually lack some aspects the trust community considers to be relevant when modeling trust. As an example, a single score value neglects information on the reliability of this very score. This reliability information is explicitly encoded in many trust models as (un)certainty [7], [10] and hidden from the user of the visualisation.

We identified several evolutionary steps in the design of trust visualisations that we discuss in the following.

<sup>1</sup>However, no significant difference was identified for human users identifying a higher and/or more reliable rating with the new stars-based visualisation in comparison with T-Viz.

### A. Stars Interface and Similar

The most popular application of communicating trust statements to human users is found in the area of e-commerce, for example, by Amazon, Inc., eBay, Inc., or Alibaba Group. Such simple visualisations use a number of symbols, mostly stars, to express a product evaluation (see Figure 1a). Here, platform operators and sellers on these respective platforms are interested in communicating a trust statement on the quality of their offered products. Sources for such trust statements are usually previous buyers, who state their level of satisfaction with the product and the buying process, as well as topic experts, who claim to provide objective assessments of a product. These statements are aggregated by the means of a trust model, often just by simple averaging, and the resulting statement is visually communicated to the prospective buyer. Thereby, the multitude of trust statements (one per product) offers a second function to the prospective buyer: comparing multiple products in order to find the one being most appropriate to his or her individual requirements.

Some of the trust community do not consider such *product quality visualisations* to be real trust visualisations as they omit some factors that distinguish a trust statement from a statement of satisfaction, like, for example, reliability metrics.

### B. Single-criterion Expert Visualisations

Early trust visualisations like the *Opinion Triangle* by Jøsang [2] were designed with broad applicability in mind. Instead of focusing on a specific application domain like e-commerce, the design goals of such *expert visualisations* are a distinct and accurate visual representation of a trust statement. An example of an Opinion Triangle can be found in Figure 1b. An opinion triangle expresses a trust statement in the notation of the *Beta Reputation System* by the same author [6], [11]. It is strongly tailored to the notation of the Beta Reputation System and includes a quintuple regarding a statement  $x$ , expressing belief  $b_x$  and disbelief  $d_x$ , as well as the associated uncertainty  $u_x$ , a base rate  $a_x$ , and an expectation value  $E_x$ . Users familiar with the Beta Reputation System can obtain all relevant information at a high level of detail. This enables experts to extensively comprehend trust statements presented

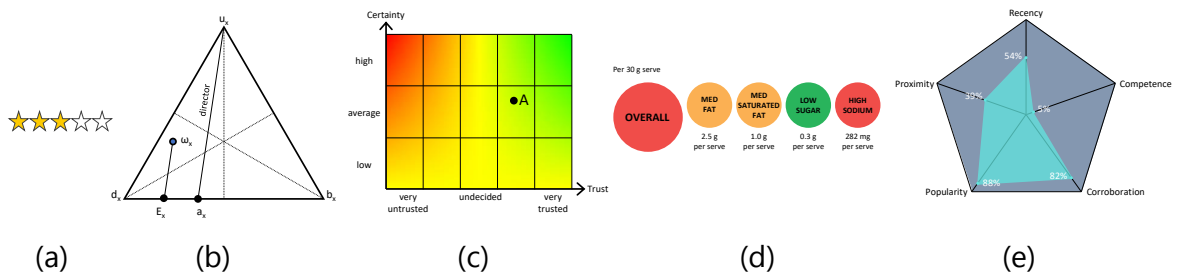


Fig. 1: Examples for trust visualisations: (a) Stars interface, (b) Single-criterion expert visualisation, (c) Single-criterion all-purpose visualisation, (d) and (e) Multi-criterion all-purpose visualisation. (Best seen in color.)

as Opinion Triangles. However, our aim in this paper is to also address non-expert users.

### C. Single-criterion All-purpose Visualisations

Instead of targeting expert users, *all-purpose visualisations* target a wider scope of users and abstract from the specific underlying trust model. An example for such a visualisation is the *Human Trust Interface* by Ries [3] (see Figure 1c). While being based on the *CertainTrust* trust model introduced in [7], the Human Trust Interface abstracts from the specific notation of CertainTrust. Therefore, it displays a trust statement in a two-dimensional grid with the axes denoting trust and certainty, respectively. Additionally, the grid is filled with a green-to-red color gradient denoting the specific *expectation value* at every point. The used visual cue is known to users from traffic lights. In [3], the Human Trust Interface is compared to the Opinion Triangle. Based on these results, the Human Trust Interface offers a clear advantage in terms of being intuitive over the Opinion Triangle.

### D. Multi-criterial All-purpose Visualisations

While the former three trust visualisations are designed to convey a single trust statement at once, multi-criterial trust visualisations address the issue of needing to have multiple trust statements on the same subject at the same time. When referring to the e-commerce example, a prospective customer often has multiple criteria that are relevant for his or her purchase decision (e.g., production quality, used materials, compatibility aspects). The naïve way to address this issue is to display multiple single-criterion trust visualisations together. However, this does not solve the issue of conveying an overall score, which could be done by adding yet another single-criterion trust visualisation.

Aside from naïve solutions as combining multiple single-criterial visualisation, multi-criterial all-purpose visualisations range from simple adaptations of traffic light-signaling, as, for example, the *front-of-pack food labelling system* investigated by Kelly et al. [12], to sophisticated visualisations like Nurse et al.'s radar plot visualisation [4], [13] and Volk et al.'s T-Viz [5]. See Figure 1d for a front-of-pack food labelling example and Figure 1e for an example for Nurse et al.'s radar plot. An example of T-Viz can be seen in Figure 2a.

## III. T-VIZ

In this section, we describe the T-Viz trust visualisation in more detail. The trust ratings represented by T-Viz consist of ratings in multiple categories and of an overall rating, which is an aggregation of all the category ratings. Each rating furthermore is assigned a reliability value. The exact formula for the calculation of this value has been the topic of research in [7], taking into account such variables as the number of positive and negative ratings. In order to represent the ratings for various specific categories and their reliability, as well as for the overall rating, T-Viz is composed as a circle chart divided into several sectors (one sector for each category). T-Viz utilizes such indicators as height, relative width, and color

of individual radar blades in each sector, as well as the centre of the chart. These indicators are designed as follows:

*a) Height:* The height of an individual radar blade represents the average rating for a particular category (higher slices represent better ratings).

*b) Width:* The width of an individual radar blade represents the reliability (in fact: the certainty) of the rating for a particular category (more relative width indicates higher certainty).

*c) Color:* The color of an individual radar blade is calculated as a combination of both the rating and its reliability, ranging from red (low average rating, very unreliable) to green (high average rating, very reliable). Thereby, the color denotes the expectation value for the given trust statement.

*d) Centre:* The centre of the chart shows the overall rating. Usually, this overall rating is calculated by applying the n-ary CertainLogic AND operator [14] to all category ratings.

T-Viz has been evaluated in two user studies, [5] and [15] and T-Viz was found to have identical decision performance as the stars interface (with a non-signification advantage of 3% for T-Viz), but decisions were made significantly faster when the study participants used T-Viz. T-Viz is used as a baseline for this paper's evaluation as it combines multiple trust statements, each of them communicating a trust value and a reliability score, and as it has advantages over presenting multiple stars interfaces.

## IV. PROTOTYPE DEVELOPMENT

The T-Viz interface is the most advanced one and it has already been evaluated with users [5], [15]. These studies show that T-Viz outperforms the standard stars interface. However, the study results indicate that there is room for even performing better than T-Viz.

We developed alternatives to T-Viz by considering findings from human-computer interaction and cognitive psychology. For this purpose, we have conducted a literature review. The result is presented in Section IV-A. The development process based on the results of this literature review as well as the proposed alternatives are described in Section IV-B.

### A. Literature Review

In order to develop our prototypes, we have conducted a literature review in the field of human-computer interaction and cognitive psychology in order to be able to integrate the psychological aspects of data visualization into the development of our trust visualization prototypes. The findings we identified as most relevant can be categorised in “graphical data representation”, “color perception”, and “comprehension”:

Research in *graphical data representation* has shown that people are more likely to comprehend the data presented in form of bar charts, if the goal of the representation was to provide a comparison (as opposed to, for example, line charts that were found to be more suitable for representing time trends) [16], [17]. Furthermore, if the data is in the form of multiple parts that together constitute a whole, circle charts

were shown to be most suitable for providing an overview on the whole and of the proportions of the individual parts [16].

Research wrt. *color perception* has shown that, although a scale from green to red is commonly used in many applications, it is not always the most appropriate color combination depending on the context. In particular, the study from [18] has researched to which extent the different colors are liked by the users (likeness), and to which extent colors urge the user to act, e.g., as a reaction to a warning or other information presented in a corresponding color (activation). The study has shown that hues of blue were most likable and hues of red and blue were the colors with the highest activation. The research in [19], furthermore, has shown that varying the saturation and brightness of color had more impact on the users' perception of the presented information than the choice of a particular hue. The work in [20] further stresses the importance of considering the color blindness of potential users.

Research wrt. *comprehension* like [21], [22] stressed the importance of familiarity: people are more likely to comprehend the data if it is presented using familiar concepts from the same context. The concept of applying common practices in developing usable interfaces has also been elaborated in [23].

## B. Prototypes

Based on the findings described in the previous section, we first designed eight preliminary prototypes for trust visualisations, which present the ratings in multiple categories, the overall rating, and the reliability of the ratings to the users. These eight prototypes were discussed in expert groups as well as with potential users. The discussions helped to identify two favorites: *Stars* and *Bars*. Examples for the Bars and Stars visualisations are shown in Figures 2b and 2c. Both are explained in more detail in the following paragraphs, including the ways the findings from Section IV-A were incorporated in the design.

1) *Bars*: The first prototype combines a bar chart with a circle chart. As the bar charts are most appropriate for comparisons, we used them in order to enable the users to compare the ratings of different products or services. Furthermore, as the circle charts have been shown to be most appropriate to represent the whole consisting of different parts, we used it in our prototype to support the users in getting the overview of all the ratings in multiple categories that influence the overall rating. The prototype also relies on the variations of brightness of the color blue to convey information, as blue has shown to be the color with the highest likeness and activation properties, and the variations of brightness were shown to be effective for influencing the perception of the presented information. The prototype uses height and width of the bar, variations of brightness of blue, and the centre of the chart as follows:

a) *Height*: The height of the bar represents the average rating for a particular category (higher is better).

b) *Width*: The width of the bar represents the total number of ratings for a specific category.

c) *Color*: The color of the bars—in particular, the brightness level of it—represents the reliability of the rating for a particular category, ranging from light blue (very unreliable) to dark blue (very reliable). For the sake of the study, we assume that reliability is calculated using the same formula and input values that are used in T-Viz (see Section III) to calculate the certainty value [7].

d) *Centre*: The centre of the chart represents an overall rating. Its color shows the average reliability over the ratings for each category (darker is more reliable), and the number shows an average rating over each category (higher is better).

2) *Stars*: The second prototype is similar to the Bars prototype with the exception of using stars instead of bars to represent data. As the star-based system has been widely established for ratings, we rely on its familiarity in supporting peoples' understanding of the ratings in our prototype. The number of stars, their color (as in Bars, variations of brightness of blue), and the centre of the chart are used as follows:

a) *Number*: The number of stars represents the average rating for a particular category, ranging from zero (very low average rating) to five (very high average rating).

b) *Color*: The color of the stars represents the reliability of the rating for a particular category, ranging from light blue (very unreliable) to dark blue (very reliable). As for the Bars prototype, we assume that reliability is computed in the same way as in T-Viz.

c) *Centre*: The centre of the chart represents an overall rating. Its color shows the average reliability over the ratings for each category (darker means more reliable), and the number shows an average rating over each category (higher is better).

## V. STUDY

We describe the user study conducted to evaluate the performance of the two proposed prototypes compared to T-Viz.

### A. Research Hypotheses

In order to evaluate the three trust visualisations, we chose to focus on three factors: correctness, efficiency, and satisfaction. *Correctness* meaning that the users would be able to correctly identify a higher or more reliable rating. In evaluating correctness, we distinguish between two cases: the first case encompasses identifying a higher or more reliable rating *in a specific category*. The second case encompasses identifying a higher or more reliable *overall rating*. The second factor, *efficiency*, is measured by the time that the users take to decide between alternatives. The third factor, *satisfaction*, describes the willingness of the users to rely on a particular trust visualisation. Note that the three factors mirror the three dimensions of usability, namely: effectiveness, efficiency, and satisfaction as defined in [24].

Hence, in order to compare our two prototypes with T-Viz, we derive the following research hypotheses:

**H<sub>1</sub>**: The three trust visualisations differ in their impact on the *correctness* of the users' perception, i.e., the ability of the users to identify a higher or more reliable rating *with regards to a specific category* differs.

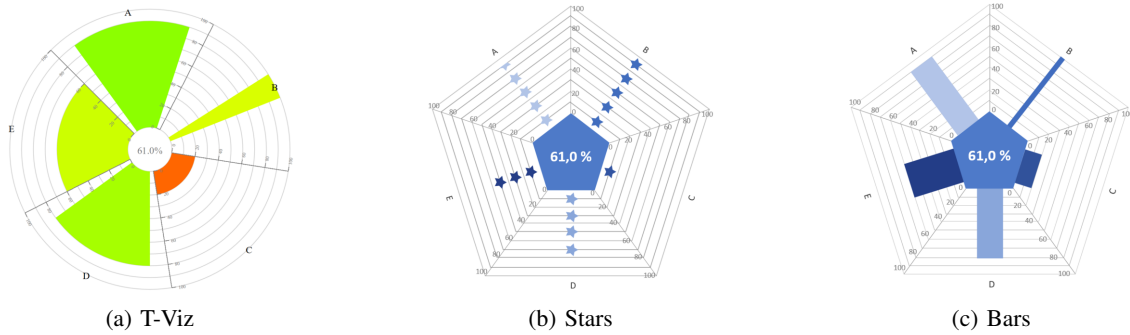


Fig. 2: Examples of the three trust visualisations.

$H_{1a}$ : The users are more likely to *correctly* identify a higher or more reliable rating in a specific category using the visualisation with Stars than using T-Viz.

$H_{1b}$ : The users are more likely to *correctly* identify a higher or more reliable rating in a specific category using the visualisation with Bars than using T-Viz.

$H_2$ : The three trust visualisations differ in their impact on the *correctness* of the users' perception, i.e., the ability of the users to identify a higher or more reliable rating with regards to an overall rating differs.

$H_{2a}$ : The users are more likely to *correctly* identify a higher or more reliable overall rating using the visualisation with Stars than using T-Viz.

$H_{2b}$ : The users are more likely to *correctly* identify a higher or more reliable overall rating using the visualisation with Bars than using T-Viz.

$H_3$ : The three trust visualisations differ in their *efficiency*, i.e., the time it takes the users to identify a higher or more reliable rating with regards to a specific category differs.

$H_{3a}$ : The users can identify a higher or more reliable rating in a specific category faster using the visualisation with Stars than using T-Viz.

$H_{3b}$ : The users can identify a higher or more reliable rating in a specific category faster using the visualisation with Bars than using T-Viz.

$H_4$ : The three trust visualisations differ in their *efficiency*, i.e., the time it takes the users to identify a higher or more reliable rating with regards to an overall rating differs.

$H_{4a}$ : The users can identify a higher or more reliable overall rating faster using the visualisation with Stars than using T-Viz.

$H_{4b}$ : The users can identify a higher or more reliable overall rating faster using the visualisation with Bars than using T-Viz.

$H_5$ : The three trust visualisations yield different *satisfaction ratings*, i.e., the users prefer a particular visualisation.

$H_{5a}$ : The users prefer using Stars over T-Viz.

$H_{5b}$ : The users prefer using Bars over T-Viz.

## B. Recruitment and Ethics

The participants were recruited via the Clickworker platform<sup>2</sup>. They were offered a reimbursement of two Euros for their participation. Ethical requirements for research involving human participants are provided by an ethics commission at Technische Universität Darmstadt.

## C. Study Design

After the welcome page, the participants were presented with a description for all the three trust visualisations, namely: Stars, Bars, and T-Viz. The order of the descriptions was chosen at random for each participant. After the participants had read the instructions, they were requested to answer two comprehension questions: 1) whether or not the rating was represented by the number of stars in Stars, by the length of the bars in Bars, and by the length of the pie segments in T-Viz. 2) Whether or not the reliability of the rating was represented by the color of the stars in Stars, by the color of the bars in Bars, and by the color of the pie segments in T-Viz. Recalling from the prototype description in Section IV-B, the correct answer to the first question is “yes”, and the correct answer to the second question is “yes” as well. The purpose of these questions was to ensure that the participants understand each trust visualisation. The participants who did not answer the questions correctly were filtered and excluded from further evaluation.

After the comprehension questions, the main part of the study started, which consisted of 12 rounds—three trust visualisations with four questions each. In every round, two images representing two different ratings with the same trust visualisation were shown to the participants. For every round, the participants were asked one of the four question types:

- 1) Which image shows a higher overall rating?
- 2) Which image shows a higher rating in category  $X$ ?
- 3) Which image shows a more reliable overall rating?
- 4) Which image shows a more reliable rating in category  $X$ ?

For each trust visualisation, every question was asked in a separate round, resulting in 12 overall rounds. The order of the presented trust visualisations and of the questions was randomized for every participant.

<sup>2</sup><https://www.clickworker.com>, last accessed on 21.03.2017

In the final part of the study, all participants were asked which trust visualisation they liked the most. As a reminder, an example for every trust visualisation was shown, presented as in Figure 2. The ratings were the same in all three examples.

## VI. RESULTS

### A. Demographics

Overall, 101 persons participated in the study. Out of those 101, eight participants were discarded from the evaluation due to the following reasons: 1) one participant did not complete the study, 2) two participants required longer than 43 minutes for completing the study; this was significantly higher (more than two standard deviations) than the median of 10 minutes and 52 seconds, 3) five participants are color-blind. In particular, the two participants who required very long to complete the study were excluded as we assumed that they were at some point distracted from the study. The color-blind participants were excluded as they were unable to rely on color, which ranges from green to red, to understanding the rating in T-Viz.

Out of the remaining 93 participants, there were 46 women and 47 men, ages 19-60 with the average age being 40.3 years and a standard deviation of 11.05 years. As for the controlled variables, e.g., age, gender etc. we did not find them to have a significant effect on the further results. Same goes for the analysis of variances for within-subject effects of the order of the visualisations ( $F(11, 1012) = 0,49, P = 0.91$ ).

### B. Research Hypotheses Evaluation

We evaluate the research hypotheses defined in Section V-A.

1) *Correctness*: First, we evaluate the hypotheses  $H_1$ ,  $H_{1a}$ ,  $H_{1b}$ , and  $H_2$ ,  $H_{2a}$ ,  $H_{2b}$ . Given a comparison of either a rating in a specific category or an overall rating for each trust visualisation, we consider three groups: 1) the participants who answered both the “higher rating” question and the “reliability” question correctly (i.e. identified both the image with a higher rating **and** the image with a more reliable rating), (2) the participants who answered either the “higher rating” question or the “reliability” question correctly (i.e. either correctly chose an image with a higher rating but making a mistake in choosing

an image with a more reliable rating or vice versa), (3) the participants who were unable to answer neither the “higher rating” question, nor the “reliability” question correctly (i.e. were unable to identify neither an image with a higher rating nor an image with a more reliable rating).

a) *Specific category –  $H_1$* : The number of participants providing a correct answer to both the “higher rating” question and the “reliability” question, to one of these questions, or to none of them with regards to a specific category are shown in Figure 3a.

The Friedman-test shows a difference in the impact on the *correctness* of the users’ decisions using the three trust visualisations regarding a specific category rating ( $\chi^2(2) = 11.29, p < .01$ ). Hence,  **$H_1$  is confirmed**.

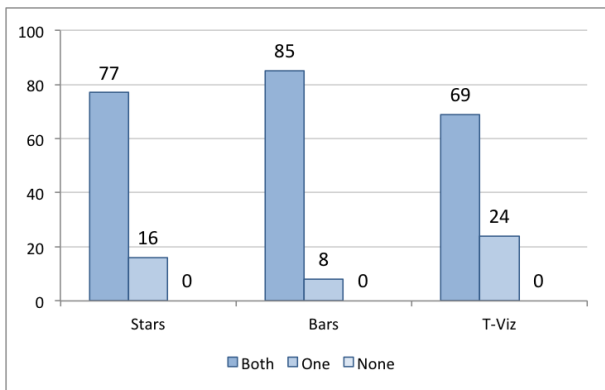
A Wilcoxon signed-rank test was conducted in order to check whether the Stars or Bars are significantly better regarding their impact on the correctness of decision making than T-Viz. The test did not show any significant difference for the comparison between Stars and T-Viz ( $V = 294.5, p = .1474$ ). Hence,  **$H_{1a}$  is not confirmed**. On the other hand, significantly more participants answered the study question correctly using Bars than using T-Viz ( $V = 218.5, p < .001$ ). Hence,  **$H_{1b}$  is confirmed**.

b) *Overall rating –  $H_2$* : The number of participants providing a correct answer to both the “higher rating” question and the “reliability” question, to one of these questions, or to none of them with regards to an overall rating are shown in Figure 3a.

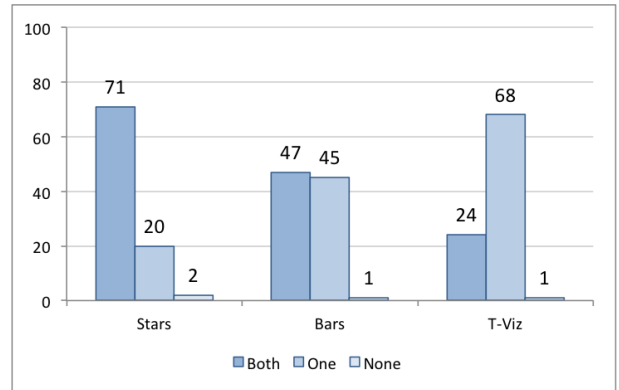
Conducting the Friedman-test reveals significant differences in the *correctness* between all the three trust visualisations regarding an overall rating ( $\chi^2(2) = 49.59, p < .001$ ). Hence,  **$H_2$  is confirmed**.

The Wilcoxon signed-rank test reveals that both Stars and Bars trust visualisations perform significantly better regarding their correctness than T-Viz, with test ratings of  $V = 1375, p < .001$  for Stars and  $V = 672, p < .001$  for Bars. Hence,  **$H_{2a}$  and  $H_{2b}$  are confirmed**.

2) *Efficiency*: We proceed with evaluating the efficiency of the trust visualisations, namely, with hypotheses  $H_3$ ,  $H_{3a}$ ,  $H_{3b}$ ,



(a) Specific category



(b) Overall rating

Fig. 3: Correct answers to the “higher rating” question and the “reliability” question for the three trust visualisations.

and  $H_4$ ,  $H_{4a}$ ,  $H_{4b}$ . For evaluating efficiency, we measured the time it took the participants to answer each question in the study and compared the averages for each visualisation method.

a) *Specific category –  $H_3$* : The average times, and the respective standard deviations, that the participants needed to answer the study questions regarding a specific category are depicted in Figure 4a.

We use the decision time as a measure for the *efficiency* of the trust visualisations. As the Mauchly test of sphericity is significant and therefore assumption of homogeneity of variance is violated, we tested this hypothesis with non-parametric tests. The result of a Friedman-test shows significant differences in the measured decision time for the trust visualisations, regarding a specific rating ( $\chi^2 = 10.21$ ,  $p < .01$ ). Hence,  **$H_3$  is confirmed.**

The Wilcoxon signed-rank test reveals that the participants were significantly faster in making their decision with both Stars and Bars than with T-Viz, with  $p < .05$ ,  $V = 1632$  for the comparison between Stars and T-Viz and  $V = 1691$ ,  $p < .1$  for the comparison between Bars and T-Viz. Hence,  **$H_{3a}$  and  $H_{3b}$  are confirmed.**

b) *Overall rating –  $H_4$* : The average times and standard deviations that the participants needed to answer the study questions regarding an overall rating are depicted in Figure 4b.

A repeated measures ANOVA has shown that there are significant differences in the *efficiency* of the three trust visualisations ( $F(2, 184) = 5.45$ ,  $p = .006$ ). Hence,  **$H_4$  is confirmed.**

The post-hoc tests also show that the participants required significantly less time for their decision using both Stars and Bars visualization methods than using T-Viz ( $p = .014$ ,  $T = -2.51$  for Stars,  $T = -3.33$ ,  $p < .001$  for Bars). Hence,  **$H_{4a}$  and  $H_{4b}$  are confirmed.**

3) *Satisfaction –  $H_5$* : We finally evaluate our hypothesis  $H_5$ , i.e., we evaluate the preference of the study participants to use one of the three trust visualisations. The number of participants that prefer each one of the three trust visualisations is depicted on Figure 5.

The chi-square goodness of fit test shows no significant difference between the number of participants that prefer each trust visualisation ( $\chi^2 = 2.32$ ,  $p = .31$ ). Hence,  **$H_5$ ,  $H_{5a}$  and  $H_{5b}$  are not confirmed.**

### C. Stars vs Bars

In addition to the hypotheses meant to compare T-Viz with our two proposed trust visualisations, Stars and Bars, we also conducted tests to compare Stars with Bars. As such, the Wilcoxon test shows significant differences between the impact of either Stars and Bars on the *correctness* of the users' decisions, both in a specific category ( $V = 34$ ,  $p < .05$ ) and on the overall score ( $V = 476$ ,  $p < .001$ ). At the same time, neither Stars nor Bars performs decisively better than the other one. As such, our participants were able to answer the questions regarding an *overall score* better using Stars than using Bars. However, Bars has a better impact on correct decisions regarding a *specific category*. The Wilcoxon test and

the paired T-test analyzing the *efficiency* does not show any significant differences among Stars and Bars, neither for a specific category ( $V = 2133$ ,  $p = .84$ ), nor for an overall rating ( $T = .28$ ,  $p = .78$ ).

## VII. DISCUSSION AND FUTURE WORK

Both the Stars and the Bars trust visualisations are shown to perform either better—or not significantly different—from T-Viz. In particular, Bars performs significantly better regarding both efficiency and correctness, for overall scores and for specific categories, while Stars does not significantly differ from T-Viz in correctness regarding an overall rating, but performs significantly better than T-Viz for efficiency and for correctness regarding a specific category. No significant difference regarding satisfaction has been revealed. The comparison between Stars and Bars shows that Bars performed better in correctness regarding a specific category, while Stars performed better in correctness regarding an overall rating. We note, however, that the presentation of the overall rating is the same in both of the prototypes, and it is possible that the instructions given to the participants influenced their decision making. Hence, we consider Bars to be more promising for further development.

Furthermore, our results did not show that the participants actually preferred using either Stars or Bars over using T-Viz. This demands further studies that investigate the factors that make T-Viz attractive, even as the participants were more capable of answering correctly to the questions using our two proposed trust visualisations, requiring less time to answer these questions.

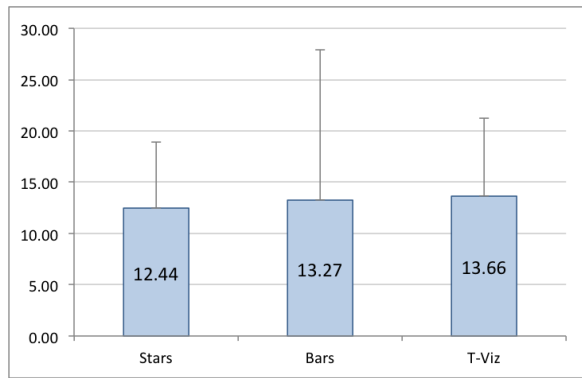
*Limitations:* Our study recruited the participants from Germany, hence, it is to be determined, to which extent the results can be transferred to other cultural contexts. It also used abstract categories instead of examples for concrete products or services. Hence, it is yet to be determined, whether there are specific types of products or services that are particularly suitable or not suitable for our proposed trust visualizations.

## VIII. CONCLUSION

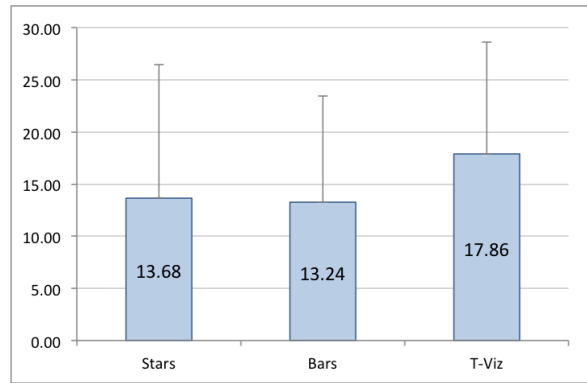
In this paper, we presented two novel trust visualisations based on the recently proposed T-Viz [5]. The visualisations replace T-Viz' "radar blades" with stars known from popular e-commerce platforms, and with bars known from bar charts, respectively. In an attempt to make trust visualisations more accessible to non-expert users and to foster wide-spread use of trust visualisations, we integrated findings from human-computer interaction and cognitive psychology into our design process.

With the help of a user study with 101 participants, we compared the correctness and efficiency when using our proposed trust visualisations with T-Viz. Results indicate that our proposals support human users better than T-Viz when identifying a higher and/or more reliable rating. Moreover, human users are quicker in reading our proposed visualisations than reading T-Viz.

Aside from these findings, we recommend further research in the field of user acceptance: in an open question for their



(a) Specific category



(b) Overall rating

Fig. 4: Average time (s) for answering the questions using the three trust visualisations.

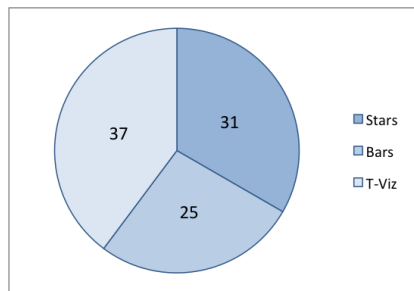


Fig. 5: Number of participants preferring each visualisation.

subjective preference, no significant preference was given to either one of the two proposed visualizations, and in fact, most study participants still named T-Viz as their preferred visualisation. Such a result is surprising, as the objective figures indicate otherwise, and as such, it deserves further investigation and will be addressed in future work.

#### ACKNOWLEDGMENT

This work has been co-funded by the DFG as part of project D.1 within the RTG 2050 “Privacy and Trust for Mobile Users”. It has also been supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts within CRISP. This paper is also supported by European Unions Horizon 2020 research and innovation programme under grant agreement No 740923, project GHOST (Safe-Guarding Home IoT Environments with Personalised Real-time Risk Control).

#### REFERENCES

- [1] L. Rasmusson and S. Jansson, “Simulated Social Control for Secure Internet Commerce,” in *NSPW*. NY, USA: ACM, 1996, pp. 18–25.
- [2] A. Jøsang, “An Algebra for Assessing Trust in Certification Chains,” in *NDSS*. The Internet Society, 1999.
- [3] S. Ries and D. Schreiber, “Evaluating User Representations for the Trustworthiness of Interaction Partners,” in *ReColl’08 in Conjunction with IUI’08*. ACM Press, 2008.
- [4] J. R. C. Nurse, I. Agrafiotis, S. Creese, M. Goldsmith, and K. Lamberts, “Communicating Trustworthiness using Radar Graphs: A Detailed Look,” in *PST*, 2013, pp. 333–339.
- [5] F. Volk, S. Hauke, D. Dieth, and M. Mühlhäuser, “Communicating and visualising multicriterial trustworthiness under uncertainty,” in *PST*, 2014, pp. 391–397.
- [6] A. Jøsang, *Subjective Logic - A Formalism for Reasoning Under Uncertainty*. Springer, 2016.
- [7] S. Ries, “Extending Bayesian Trust Models Regarding Context-Dependence and User Friendly Representation,” in *ACM SAC*, 2009, pp. 1294–1301.
- [8] V. Kant and P. Dwivedi, “An evidential trust model for web services based on fuzzy sets,” *Procedia Computer Science*, vol. 57, pp. 537–544, 2015.
- [9] S. M. Habib, F. Volk, S. Hauke, and M. Mühlhäuser, “Computational trust methods for security quantification in the cloud ecosystem,” *The Cloud Security Ecosystem*, pp. 463–493, Jun. 2015.
- [10] A. Jøsang, “A Logic for Uncertain Probabilities,” *Intl. J. of Uncert., Fuzziness and Knowledge-Based Sys.*, vol. 9, no. 3, pp. 279–311, 2001.
- [11] A. Jøsang and R. Ismail, “The beta reputation system,” in *In Proceedings of the 15th Bled Electronic Commerce Conference*, 2002.
- [12] B. Kelly, C. Hughes, K. Chapman, J. C.-Y. Louie, H. Dixon, J. Crawford, L. King, M. Daube, and T. Slevin, “Consumer testing of the acceptability and effectiveness of front-of-pack food labelling systems for the Australian grocery market,” *Health Promotion International*, vol. 24, no. 2, pp. 120–129, 2009.
- [13] J. R. C. Nurse, I. Agrafiotis, M. Goldsmith, S. Creese, and K. Lamberts, “Two sides of the coin: measuring and communicating the trustworthiness of online information,” *J. of TM.*, vol. 1, no. 1, pp. 1–20, 2014.
- [14] S. Ries, S. M. Habib, V. Varadharajan, and M. Mühlhäuser, “Certainlogic: A logic for modeling trust and uncertainty (short paper),” in *TRUST*. Springer, 2011, pp. 254–261.
- [15] F. Volk, “Detailing reviews and ratings for trust-enhanced composition,” Ph.D. dissertation, Technische Universität Darmstadt, 2015.
- [16] P. Shah and J. Hoeffner, “Review of graph comprehension research: Implications for instruction,” *Educational Psychology Review*, vol. 14, no. 1, pp. 47–69, 2002.
- [17] B. Strobel, S. Sass, M. A. Lindner, and O. Köller, “Do graph readers prefer the graph type most suited to a given task? insights from eye tracking,” *Journal of Eye Movement Research*, vol. 9, no. 4, 2016.
- [18] R. Mehta and R. J. Zhu, “Blue or red? exploring the effect of color on cognitive task performances,” *Science*, vol. 323, no. 5918, pp. 1226–1229, 2009.
- [19] L. I. Labrecque and G. R. Milne, “Exciting red and competent blue: the importance of color in marketing,” *Journal of the Academy of Marketing Science*, vol. 40, no. 5, pp. 711–727, 2012.
- [20] J. Birch, “Worldwide prevalence of red-green color deficiency,” *JOSA A*, vol. 29, no. 3, pp. 313–320, 2012.
- [21] C. Ware, *Information visualization: perception for design*, 2012.
- [22] R. J. Sternberg and K. Sternberg, *Cognitive psychology*, 2016.
- [23] T. Paul, M. Stopczynski, D. Puscher, M. Volkamer, and T. Strufe, “C4ps-helping facebookers manage their privacy settings,” in *International Conference on Social Informatics*. Springer, 2012, pp. 188–201.
- [24] Intl. Org. for Standardization, “ISO/DIS 9241-11.2. Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts.”