

Communicating Trustworthiness using Radar Graphs: A Detailed Look

Jason R. C. Nurse^{†§}, Ioannis Agraftotis[†], Sadie Creese[†], Michael Goldsmith[†] and Koen Lamberts[§]

[†] Cyber Security Centre, Department of Computer Science, University of Oxford, UK

[§] Department of Psychology, University of Warwick, UK

{jason.nurse, ioannis.agraftotis, sadie.creese, michael.goldsmith}@cs.ox.ac.uk, k.lamberts@warwick.ac.uk

Abstract—The amount of trust we, as human-beings, place in each other or an object (e.g., online information) is typically guided by several trust factors and antecedents. These factors can vary in importance depending on the individual making the trust decision and also on the situation – such is actually the subjective nature of trust. In this paper, we explore this notion of factors’ importance by delving into detail on some of our recent user experiments and subsequent findings, partly described in previous work. These experiments used radar graphs to communicate trustworthiness as a function of five trust factors, namely competence, popularity, recency, corroboration and proximity. Here, we expand that work by further considering the importance of each of the factors to participants, while also investigating the correlations between individuals’ perceptions of trust, and aspects such as graph area or size and expected scores as calculated by linear regression analysis. More specifically, we focus on outliers and endeavour to understand what is the cause of their existence. This research contributes to the field of communicating trustworthiness now, but is also meant to act as a platform for future, more directed research on visuals intended to communicate trustworthiness.

Keywords—Trustworthiness factors; trustworthiness visualisation; risk communication; communicating trustworthiness and quality; user studies

I. INTRODUCTION

Trust is an essential part of modern-day interactions, featuring in everything from physical business transactions to online health advice and e-commerce. As a result of this significance, several researchers (e.g., [1–6]) have investigated the nature of trust in depth, with the aim of facilitating a better understanding and appreciation of this social concept. The outcome of these articles has been the proposal of numerous novel models for trust and the identification of several factors and antecedents which influence and guide people’s trust decisions. In many ways, these factors make trust much more tangible and act to provide an idea for when something or someone might or might not be trusted.

In this paper, we extend some of our recent user-study work in [7] which focused on two aspects. Firstly, assessing the importance of five factors, i.e., Competence, Popularity, Recency, Corroboration and Proximity, as they pertained to trust; the potential for automated assessment being the reason behind the choice of these specific factors. Secondly, the utility of those and similar factors (displayed in a radar graph) as a tool to build confidence in a single information-trustworthiness metric (visualised as a traffic light). The overall

goal of this metric was to support decision-makers using online information in distinguishing to what extent to trust content and its source; articles with similar general aims can be seen in [8–10].

To build on that work, our assessment now considers the factor-importance findings from [7] in more detail and also investigates the correlations between participants’ perceptions of trust and features such as graph size and expected scores (which are calculated based on the trustworthiness equation found in [7]). We particularly focus on the outliers (i.e., cases where there was significant deviation from the normal or anticipated participant perception or behaviour), aiming to understand the cause of their existence. To that end, we take into account the possible explanations described above. We examine whether alteration in participants’ perception of trust was a result of the graph’s size, of the time constraint or of a different ordering on the importance of the five trustworthiness factors. We also assess how well the trustworthiness formula (from [7]), actually captures participants’ scores. Most importantly, we view the findings from this current paper as an ideal springboard to further and more directed research on graphs as an information-trustworthiness visual, and as such dedicate an entire section to this in the paper. This work will undoubtedly benefit from regular reflections in the risk-communication field, and the contributions in the visual communication of risks (e.g., [11–13]) and limitations of human perception and working memory ([14,15] being particularly pertinent articles).

In the remainder of this paper, we present related work on trust factors and approaches to communicate trustworthiness and more generally, risk, in Section II. Section III recaps our recent work in [7] to give more context to the research detailed in this paper. Next, Section IV presents the additional analysis conducted on the importance of trust factors and the use of radar graphs as an information-trustworthiness communication tool. In Section V, we then highlight research avenues we intend to pursue to further the understanding of factor importance levels and trustworthiness visuals, before concluding the paper in Section VI.

II. RELATED WORK

There have been many attempts in the literature to characterise and model information trustworthiness and the factors that influence it. One such article is [1], where the authors engage in a comprehensive survey of the field and definition

of an integrated model for trust. The key novelty of their model is its comprehensiveness, spanning from preconditions of trust (such as, potential harm or need to make a decision), to the factors that constitute and influence trustworthiness (e.g., objectivity, accuracy and validity) and the trust development and reinforcement process itself.

In addition to models for trust, there has also been work – most notably in [16] – towards the definition of a variety of factors, which influence it, particularly from an online perspective. Of most relevance to our work however, is their definition of four core factors that may be especially important in making trust decisions, namely, an information source’s authority, the related sources supporting the information, its provenance, and whether or not there is any bias. This therefore starts to introduce the idea of the varied importance of factors. Another article which considers the notion of trust-factor importance is [17]. There, the researchers conduct an insightful evaluation of the importance of a number of trust factors, such as reputation, look and feel, and third-party seals, and demonstrate that factor’s importance levels can vary across different contexts; contexts even as close as e-banking and e-commerce. These are all key articles which have fed into our research on trust factors in this paper and in our own prior review of the fields of information trustworthiness and quality [18].

As the amount of information online increases, there is a growing need to provide (automated) evaluations of content, which can help users to know the extent to which they can trust information. Consequently, researchers have sought to create frameworks and metrics for trust [9,19,20] to fulfil these requirements and support decision-making online. Along with the actual information-trustworthiness measurements, there has also been some work on how these values should be communicated to users. In [8] for example, a simple traffic-light system (with red, amber and green) is used to convey quality; a technique we also effectively apply to trust in [10]. While in [9], the authors take a more fine-grained approach to presenting trust as they colour the text background of Wikipedia content – from white (high trustworthiness) to dark orange (low trustworthiness) dependent on how trustworthy their system deems that segment of information to be.

Work in [21] also focuses heavily on visuals, particularly charts and graphs, as a support tool to assisting information users assess the quality of content. Although useful, in general, these articles only consider trust at the higher level and seldom communicate to users the details of how the trust value was arrived at. Only in [21] is the process transparent and quality factors are identified, but as those authors note, even there, there is the potential for factor visuals to be misinterpreted or their significance lost. Moreover, there is the question of whether always presenting individuals with that level of detail may just induce information overload.

The field of risk communication is of interest to our work as well. Its relevance is linked to our aim of presenting individuals with trustworthiness data, which in effect, portrays the risk of believing and acting on information. Guidance from

this field can be found in best practices to communicate risk information, such as in [13], but also specific advice in risk communication using visuals as outlined in [11, 12]. Most notably, as we aim to use graphs to present trust factors, the latter two articles are key as we will need to be mindful of the various biases and processing limitations faced by humans. An example of such limitations may be the number of factors people might be capable of taking in easily, and the importance of using explanations and captions for risk content. In some of our other work [22], we have considered these issues, and can undoubtedly draw from those findings as well. In the next section, we overview the experiment conducted to set the context for this paper’s contributions.

III. EXPERIMENT CONTEXT

The aim of our radar graphs experiment (initially presented in [7]) was to explore people’s perception of five trust factors, and also assess their importance to individuals as they pertain to judgements on trustworthiness. These factors were: Competence (*Cm*), the level of knowledge of a person or information source; Proximity (*Pr*), the geographical closeness of a source to an event of interest; Popularity (*Po*), how well-known is a source; Recency (*Re*), how recent or up-to-date is information to the event of interest; and Corroboration (*Cr*), how well supported is information by a variety of different sources. The experiment design consisted of 200 radar graphs, each presenting ranges of values between 0-100% for the five factors; Excel’s *RAND()* function was used to produce a set of 200 random values which was then implemented in our application for experimentation. Figure 1 shows an example visual, which displays Graph #22; with Recency at 54%, Competence at 5%, Corroboration at 82%, Popularity at 88%, and Proximity at 39%. Throughout this paper we present graphs exactly as they were presented to study participants to give readers the best insight into the experiment’s conditions and tasks.

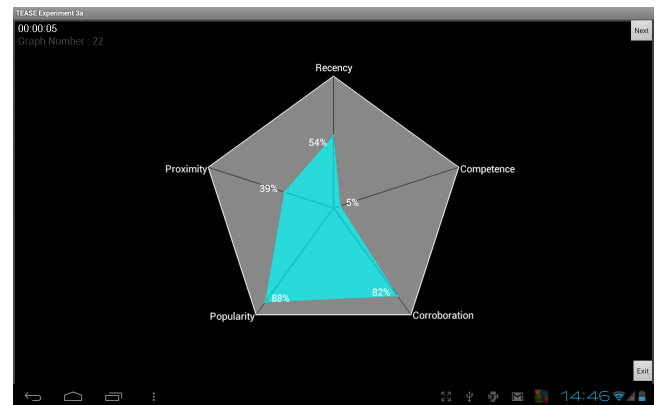


Fig. 1. Screenshot of graph app showing Graph #22

A total of 40 individuals (29 females, 11 males, mean age of 23.725, age range: 18-58 years) participated in the study.

Recruitment was conducted through the use of flyers posted within the University of Warwick and University of Oxford. Participants were from a variety of disciplines (science, arts, and social sciences) and there was also diversity of levels, i.e., students were both postgrads and undergrads, and working professionals spanned from hospitality clerks to personal assistants, researchers and administrators. Participants were compensated for assisting with the experiment's sessions.

The experiment consisted of participants being presented with each of the 200 graphs and then given a maximum of 10 seconds before they were asked to give a rating of 0-100 to represent the level of trustworthiness the graph conveyed to them. A timer was displayed on screen and therefore participants were always aware of the time remaining. The restricted time allowed served two purposes. Firstly, we were aiming to get their first and instinctive impression, and secondly, we would decrease the chances of study participants recalling how they assessed similarly shaped graphs, thereby avoiding them simply using their memory.

To present the graphs to participants, we used a Motorola Xoom tablet PC. At the beginning of the experiment, participants were briefed on the goals of the study and requested to sign a consent form. To ensure that they had a clear understanding of the five trust factors, they were also shown short definitions and examples of how the terms could be used. As there were several graphs, participants were advised that in case of discomfort (e.g., tired eyes), they were free to take a break at any time. Regarding the responses to the graphs, it was emphasised that there were no correct or incorrect answers. Once participants were comfortable, the experiment commenced, and they were asked to evaluate the trustworthiness degree represented by each of the 200 radar graphs by assessing measures of the five factors included and any personal preferences they held.

The findings from the analysis as reported in [7] were very encouraging and highlighted distinct significance levels of factors across participants. Specifically, we used linear regression analysis to identify importance (via coefficients) for each factor per participant, and then averaged across the sample to define a regression formula for the group of participants. The formula for trustworthiness (as a function of the five factors) that resulted is shown below:

$$\text{Trustworthiness} = -5.425 + 0.176Re + 0.405Cm + 0.235Cr + 0.127Po + 0.141Pr \quad (1)$$

Coefficients preceding each factor were taken to define factor importance. Thus, Competence was the most influential factor, followed by Corroboration, Recency, Proximity and Popularity. Finally, it is worth mentioning that participants did not report any significant difficulties in understanding the factors (or their relation to trustworthiness), graphs or combining them to deduce an overall trust score. This built our confidence in the findings above.

IV. GRAPH ANALYSIS

In addition to the more general evaluation in [7], we have subsequently engaged in several smaller and more focused statistical analyses pertaining to graphs and respective participants' scores. In the first investigation, we conducted a basic relationship analysis to verify that values produced by the sample's formula above (hereafter, 'expected values') correlated with participants' trustworthiness scores for each of the 200 graphs. This was to establish some link between expected values and participants' actual scores rather than an influence or causal association – although both types of values would have the underlying trust-factor scores as a basis. Another key benefit was also the potential to identify any outlying participants (i.e., those with different opinions than the general populous) that might have been marginalised after averaging across coefficients to define the sample's trustworthiness formula.

The analysis consisted of first computing expected values (using the formula in (1) and respective factor scores) for all graphs, and running a Pearson product-moment correlation [23] to determine the relationship between these values and each of the 40 participants' graph scores. As an example, in Table I, we present the details of two graphs, their respective trust-factor scores, the expected value for the graphs and the actual scores given by Participants #1 (P1) and #2 (P2). From the analysis conducted, we found a positive correlation between the expected and actual scores, with all Pearson correlation coefficients statistically significant at $p < 0.001$. This confirmed the link between these two values and suggested that there were no extreme outliers (i.e., participants with very contrary opinions and perceptions of factor importance).

Graph #	Re	Cm	Cr	Po	Pr	Expected	P1	P2
22	54	5	82	88	39	42	40	40
33	1	14	12	30	9	8	9	5

TABLE I
GRAPH DATA, EXPECTED VALUES AND SCORES FOR TWO PARTICIPANTS

We did, however, identify a participant with a Pearson coefficient of 0.353 which was much smaller than the sample's average of 0.745. Upon further investigation, we found that this participant viewed the importance of trust factors very differently, with Recency being most important (0.385), followed by Proximity (0.184), Corroboration (0.162), Competence (0.135) and then Popularity (0.081). This variation could therefore explain the lower Pearson coefficient. Generally however, the strong correlation in other scores did emphasise the utility of the main formula presented in (1).

To give further insight into how well the formula satisfied the range of participants' scores, we assessed the percentage difference between the expected values and each participant's graph score. The results from this assessment verified those above as we found that most participants gave scores reasonably close to the expected value; this also further supported the sample formula for trustworthiness. We did, however, identify a set of graphs that were persistent outliers where

some participants allocated much higher or lower scores than expected. Our analysis therefore focuses on understanding the cause of these outliers (i.e. graphs that the expected and actual score differ more than 50%) in an attempt to comprehend their nature and, if possible, mitigate their future effects.

Elaborating on the outliers, we noted that in 4 graphs, at least 40% of participants agreed in their reporting of scores greater than 50% different to expectations. Specifically: for Graph #3, 53% of participants provided scores 50% or more greater than the expected value; for Graph #45, 45% of individuals gave scores of at least 50% smaller; for Graph #52, 40% of study participants supplied scores of at least 50% greater than expected; and for Graph #66, 45% of participants reported scores of 50% or more smaller than the expected values. We noted two potential reasons for these significant variations. Either, there was a marked deviation in factors' importance (i.e., coefficients) for those groups of participants, or possibly the overall size or area of the radar graph (i.e., the filled part of the pentagon) or indeed, its schema, influenced individuals. An example of the latter case is that smaller-sized graphs, independent of specific factor importance levels, subconsciously swayed participants to give lower scores than normal for the sample.

To ascertain which of these (if any) may have been true, we first calculated the average coefficient value for each trust factor, for the four groups of participants that gave the differing scores to the graphs. One should recall that coefficients resulted from linear regression analysis on each participant's 200 graph scores. Next, we compared these coefficients (i.e., levels of importance) to those of the main sample. This assessment highlighted several variations in factors' importance, to the extent that it could have been the reason for differing final graph scores.

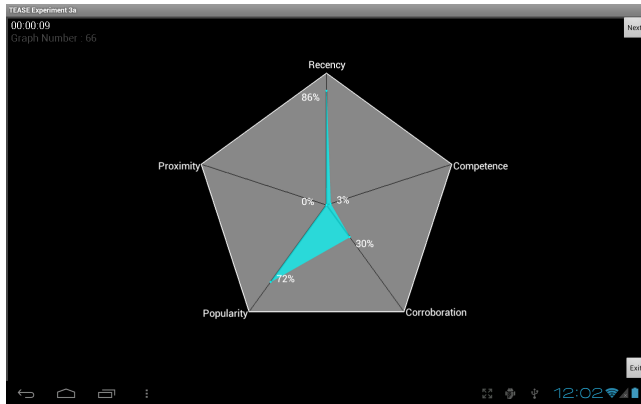


Fig. 2. Screenshot of Graph #66

In Graph #66 (in Figure 2) for instance, where 45% of participants gave scores 50% or more lower than expected, we found that these participants felt that Recency and Popularity were respectively 11% and 17% less important than in the overall sample formula and Corroboration was 9% more important. Reflecting on the graph itself, one can begin to understand

why a lower score from those participants might therefore have resulted. That is, the factors with higher graph values were less important and those with lower values (here, Corroboration only) were more important. Similar observations were apparent with the other three graphs. This was a notable finding that pointed to deviation in factors' importance as the actual reason for different (or outlier) scores, and potentially not subconscious influences of graph size. A possible solution to mitigate the effect of this specific outlier could be to allow users to customise the importance of the five factors according to their trust perceptions.

Apart from comparing participants' scores to the expected values for the sample, another more focused approach to determine whether size of graphs subconsciously influenced participants was to compare their individual scores to their expected values (i.e., values calculated based on their own trustworthiness formulae). This was therefore moving away from assessing outliers as it pertains to the sample's expected score, to identifying and evaluating outliers relating to the respective participant's expected graph score. For this investigation therefore, we used the formula generated for each participant based on the linear regression analysis, and compared the new expected values to the actual scores that were given to the respective graphs. To gather an idea of whether graph size or area might have influenced participants, we checked for cases where actual scores were 50% or more different to the expected scores, *and* size of the graph (or more accurately, the calculated area of the filled graph) was at least 25% different to the expected value. To calculate the filled area, we used the equation presented below in (2). This splits the pentagon into five triangles, calculates the blue-shaded area in each, sums these, then divides by the total pentagon area to determine the size percentage filled.

$$\frac{\frac{1}{2} \times \sin 72^\circ \times [Re \times Cm + Cm \times Cr + Cr \times Po + Po \times Pr + Pr \times Re]}{\frac{5}{2} \times 100 \times 100 \times \sin 72^\circ} \quad (2)$$

Therefore, if a participant gave an actual graph score 65% greater than their expected value for that graph, and the graph size is 40% bigger than that expected score, we hypothesised this to mean that size may have had some impact on their decision to award that higher graph score.

The results from this analysis indicated that there were a significant number of cases – i.e., 49% of the cases where there was larger than a 50% difference – where this situation occurred, which led us to believe that graph size may have had some noteworthy impact on participants' scores. Considering the situations where this transpired, we became interested in the respective graphs and whether a particular type of schema (i.e., graph arrangement) or area may have resulted in under or overestimations in scores. This could give us valuable further insight into how graphs were perceived by participants in addition to allowing us to look in more detail at the outliers present.

In testing for this, we found that for a number of the cases, particularly small sizes did feature. The graphs most

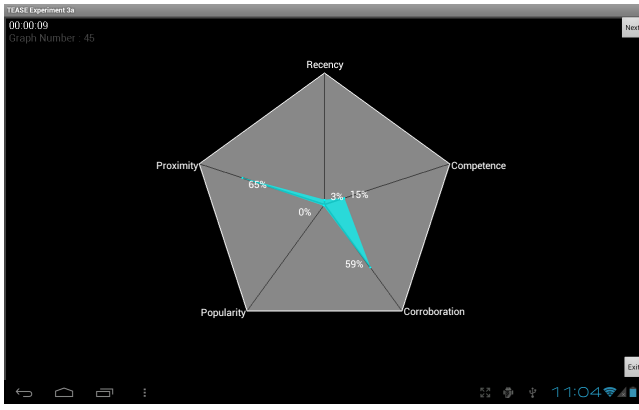


Fig. 3. Screenshot of Graph #45

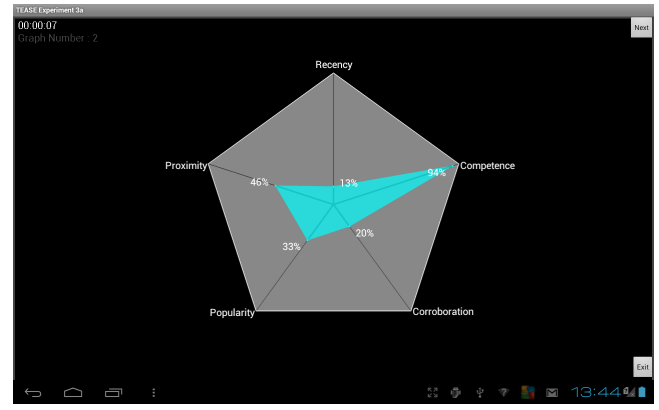


Fig. 4. Screenshot of Graph #2

underestimated by participants, and therefore those where size may have had a real influence, were Graph #45 shown in Figure 3 and Graph #66 previously presented in Figure 2; 17 participants underestimated these graphs. Assessing Graph #45's schema, the very low values for Recency, Popularity and Competence led to a particularly small graph size with considerably thin filled areas for even the higher rated factors of Proximity and Corroboration. Graph #66 exhibited a similar size phenomenon just with different trust factors. The other graphs which may have been particularly influenced by graph area are documented in Table II; incidentally, all of these graphs result in underestimations of trustworthiness scores. Readers can easily recreate the graphs to view their schemas as necessary.

Graph #	Re	Cm	Cr	Po	Pr	No. affected by size
2	13	94	20	33	46	9
15	21	0	31	70	15	9
33	1	14	12	30	9	11
106	19	14	30	0	32	10
143	63	0	2	57	37	10
148	63	2	25	25	1	11
150	35	1	73	35	7	9

TABLE II
GRAPHS WHERE SIZE MAY HAVE IMPACTED SCORES

Further to the test above, we also considered the possibility that extremely low levels of Competence (generally the most important factor) may have been the cause of low scores across the participants. This would be easily conceivable as a low score in the most important factor may have led to individuals completely discounting the graph. As can be seen in Graph #2 in Table II and Figure 4 however, even in cases where Competence was extremely high, there was some level of understatement in participants' scores. The outliers identified in this case will require further research is required to fully characterise and mitigate their effects.

The final cause that we considered as a potential source of outliers was the time constraint of 10 seconds that participants had in which to provide their answers for each graph. We

observed, however, that all the participants would respond within the first five to seven seconds, thus rendering the effect of time probably insignificant.

V. NEXT STEPS

There are several avenues for future work which we intend to pursue, in order to follow up on the analyses and exploratory findings in this paper. The first pertains to the layout of the graph itself – i.e., the position and ordering of each trust factor and its axis – and what happens when the positioning or ordering is changed. To take Graph #66 as an example, if Recency (86%), Popularity (72%) and Corroboration (30%) were next to each other, the graph's schema would be noticeably different and the area would jump from 5% to 17%. Figure 5 shows a mocked-up example of this altered format.

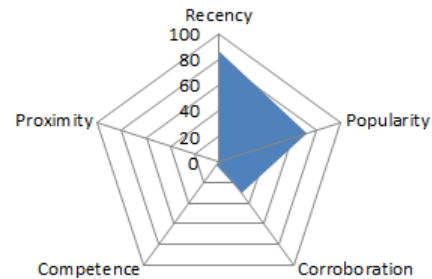


Fig. 5. Graph #66 with different factor ordering

The research question therefore remains, might these variations in ordering have an impact on participants' perceptions and the scores that they award? Or, do people see past the area and schema differences and award similar scores regardless of ordering? If ordering does have an impact then we would need to consider whether there are any 'better' orderings or indeed, whether the choice of ordering should be left to system users to decide. Finding the optimal ordering solution may mitigate the outlier effect caused by the graph size as well, thereby potentially rendering future attempts to communicate

trustworthiness as effective instead of problematic, as they seem to be now.

We have already started to consider orderings that would maximise graph size where the graph might be unduly underestimated by some individuals because of its schema. Preliminary results suggest that a graph factors ordered in the following way may result in a consistently maximised graph area: largest factor, second largest factor, forth largest factor, fifth largest factor and then third largest factor (i.e., 1-2-4-5-3). We need to be careful however, because this approach could lead to consistent overestimations in the trustworthiness of graphs and related information, since the participants will always view the maximum volume of the graphs. Moreover, users might become confused with the constant change of the trustworthiness factors in the edges of the pentagon; usability will also therefore be an issue.

In future work we could assess whether placing the trust factors with the same ordering in the radar graph (i.e., 1-2-4-5-3) but instead of prioritising them according to the highest score, the focus would be on the importance of the trustworthiness factors to the participants. This approach allows the system to provide higher volume in graphs that should be trusted and lower volumes in graphs that should not be trusted. Unfortunately, a side effect of this is that it leads to system biases, not to mention the initial task of ascertaining how important individual factors are to users (although, this could arguably be expressed by them at system setup). Nonetheless, as mentioned above, continual work on this and other approaches will be necessary.

Another avenue for further work picks up on a feedback point from interviews with participants that have undertaken the graphs experiment. There were suggestions that we should investigate the potential of additional axes in the radar graph to represent more factors. Our current emphasis on 5 factors was based on simplicity and reducing cognitive effort required by system users (a guideline from the risk communication and usability fields), but as several research articles have suggested (e.g., [14, 15]), the human brain may be capable of coping with possibly 7 or 9 items. Future experiments might therefore seek to evaluate people's ability and desire to assess additional trust factors – of which there are many [18] – and assess whether there is as preferred number of factors that should be displayed. This could also allow us to further validate our existing work and investigate the importance of other factors as it pertains to trustworthiness. Dependent on the research available, we may need to conduct a broader study on factors' importance outside of graphs and then use a subset (e.g., most important) for display within graphs during decision-making. In terms of graphs, the field of risk communication (especially seminal research work such as [11, 12]) will undoubtedly continue to be a key area of reference during our study design and subsequent user experimentation.

Departing from the increase in trust factors, an alternative approach to graphs that could illustrate multiple dimensions in a schema with clarity could also be considered (e.g., 3-D representation of a multiple-dimension schema). Results from the

new approach could be compared with results obtained when the information is conveyed by radar graphs to unveil biases, flaws and advantages that different visualisation techniques (such as graphs, histograms and 3-D schemas) may have. In a similar vein, additional experiments with comparative graph representations could provide further insights into how the shaded area of schemas may dictate or influence decisions. A logical and very practical follow-on step would be to present multiple graphs at one time (e.g., side-by-side), where the trust factors would be shown in different orderings, thus allowing users to determine what suits their preferences best.

An area where we may also conduct further research is to consider the impact of *context* on individuals, and assess the influence it has on people's perception of trust factors and their importance both within and outside of visuals and graphs. For the previous experiments, we specifically avoided context as a variable because we were more focused on how people perceived the importance of those factors generally. In the real-world however, as has been highlighted in existing research [1, 16], context is crucial and is likely to have a notable influence on factors' importance and perceived trust. This was also hinted at in interviews with participants as some said that they may have allocated different scores to the same graphs if they were presented with different scenarios. For example, in an emergency event (i.e. earthquake) the trustworthiness factors of recency and proximity could become more significant than competence.

Our future work will therefore focus on context, and evaluate if and how perception of graphs and significance levels of factors may vary dependent on context. We envisage defining context using a number of scenarios. For example, assuming that the information (to which this trustworthiness graph pertains) is to be used to determine whether to buy a new smartphone, what is the trustworthiness of the graph? Or, what if the information was to be used to determine whether or not to administer a home-made remedy to an ailing friend, how trustworthy would you rate the graph? It will also be intriguing to analyse whether there are any common or majority views across individuals that could lead to a convergence in factors' importance levels within specific contexts.

Finally, we aim to assess the impact that cognitive biases may have on the perception of graphs within our experiments, and in the context of real-world scenarios. Cognitive biases refer to systematic weaknesses in human's cognitive processing and have been discussed at length in several articles [24–26]. The bias of most immediate interest is the anchoring effect, which defines the tendency of decision makers to systematically base judgements on initial (and potentially even irrelevant) information [26]; future decisions are 'anchored' or biased to that starting information. We would therefore be specifically investigating (using quantitative – i.e., scores – and interviews approaches) whether such an effect is prevalent when individuals score graphs, and if it is, to what extent. Key questions include, assessing whether we could find clear links between this and incorrect perception of certain graphs. Furthermore, if it does exist, how does it feature in real-

life scenarios (where graphs are not presented in such quick succession and there is more situational context), and are there any techniques that could be applied to reduce its effects? A good reference point for this information is [27] and their recent, comprehensive review on the anchoring effect, its causes and attempts to tackle it in the past.

VI. CONCLUSION

As researchers seek to support decision-making online by providing information-trustworthiness indicators, the importance of understanding exactly what information is to be communicated to users, and how, increases dramatically. In this paper, we have taken steps towards understanding how to communicate trustworthiness to individuals using radar graphs as a visual tool. This work expands on our recent research by engaging in a detailed consideration of the perceived importance of five trust factors presented using radar graphs. In addition to demonstrating the importance of the factors at a sub-group and individual level (e.g., to explain lower Pearson coefficients), we also showed that our formula adequately captures and models a majority of participants' scores.

Moreover, we found some evidence to suggest that in some cases, the size and schema of a graph – independent of factors' importance – may have affected its perceived trustworthiness. This is interesting as it could highlight scenarios where graphs may be misinterpreted and therefore could result in an inaccurate user decision being made. In future work, we will assess these and other issues, with the aim of facilitating better, and more effective communication of trustworthiness online.

ACKNOWLEDGMENTS

We would like to thank HW Communications Ltd for implementing the tablet PC prototypes used in the experiments. More broadly, this work was conducted as a part of the TEASE project, a collaboration between the University of Oxford, University of Warwick, HW Communications Ltd and Thales UK Research and Technology. The project is supported by the UK Technology Strategy Board's Trusted Services Competition (www.innovateuk.org) and the Research Councils UK Digital Economy Programme (www.rcuk.ac.uk/digitaleconomy).

REFERENCES

- [1] K. Kelton, K. R. Fleischmann, and W. A. Wallace, "Trust in digital information," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 3, pp. 363–374, 2008.
- [2] A. J. Pickard, P. Gannon-Leary, and L. Coventry, "Trust in 'E': Users' trust in information resources in the web environment," in *ENTERprise Information Systems*, ser. Communications in Computer and Information Science, J. Quintela Varajao, M. Cruz-Cunha, G. Putnik, and A. Trigo, Eds., 2010, vol. 110, pp. 305–314.
- [3] T. Grandison and M. Sloman, "A survey of trust in internet applications," *IEEE Communications Surveys & Tutorials*, pp. 2–16, 2000.
- [4] L.-Y. Pan and J.-S. Chiou, "How much can you trust online information? cues for perceived trustworthiness of consumer-generated online information," *Journal of Interactive Marketing*, vol. 25, no. 2, pp. 67–74, 2011.
- [5] G. L. Urban, C. Amyx, and A. Lorenzon, "Online trust: State of the art, new frontiers, and research potential," *Journal of Interactive Marketing*, vol. 23, no. 2, pp. 179–190, 2009.

- [6] S. Marsh, A. Basu, and N. Dwyer, "Rendering unto csar the things that are csar's: Complex trust models and human understanding," in *Trust Management VI*, ser. IFIP Advances in Information and Communication Technology, T. Dimitrakos, R. Moona, D. Patel, and D. McKnight, Eds. Springer Berlin Heidelberg, 2012, vol. 374, pp. 191–200.
- [7] J. R. C. Nurse, I. Agraftiotis, S. Creese, M. Goldsmith, and K. Lamberts, "Building confidence in information-trustworthiness metrics for decision support," in *12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom-13)*, 2013.
- [8] N. Idris, M. Jackson, and R. Abrahart, "Colour coded traffic light labeling: A visual quality indicator to communicate credibility in map mash-up applications," in *International Conference on Humanities, Social Sciences, Science & Technology*, 2011.
- [9] B. T. Adler, K. Chatterjee, L. de Alfaro, M. Faella, I. Pye, and V. Raman, "Assigning trust to Wikipedia content," in *4th International Symposium on Wikis*, 2008.
- [10] J. R. C. Nurse, S. Creese, M. Goldsmith, and K. Lamberts, "Using information trustworthiness advice in decision-making," in *Workshop on Socio-Technical Aspects in Security and Trust (STAST) at 25th IEEE Computer Security Foundations Symposium (CSF)*. IEEE, 2012.
- [11] I. Lipkus and J. Hollands, "The visual communication of risk," *JNCI Monographs*, no. 25, pp. 149–163, 1999.
- [12] I. Lipkus, "Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations," *Medical Decision Making*, vol. 27, no. 5, pp. 696–713, 2007.
- [13] V. H. Visschers, R. M. Meertens, W. W. Passchier, and N. N. De Vries, "Probability information in risk communication: a review of the research literature," *Risk Analysis*, vol. 29, no. 2, pp. 267–287, 2009.
- [14] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information," *Psychological review*, vol. 63, no. 2, p. 81, 1956.
- [15] T. Saaty and M. Ozdemir, "Why the magic number seven plus or minus two," *Mathematical and Computer Modelling*, vol. 38, no. 3–4, pp. 233–244, 2003.
- [16] Y. Gil and D. Artz, "Towards content trust of web resources," *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 5, no. 4, pp. 227–239, 2007.
- [17] E. Costante, J. den Hartog, and M. Petkovic, "On-line trust perception: What really matters," in *Workshop on Socio-Technical Aspects in Security and Trust (STAST) at 5th International Conference on Network and System Security (NSS)*. IEEE, 2011, pp. 52–59.
- [18] J. R. C. Nurse, S. S. Rahman, S. Creese, M. Goldsmith, and K. Lamberts, "Information quality and trustworthiness: A topical state-of-the-art review," in *International Conference on Computer Applications & Network Security*. IEEE, 2011.
- [19] A. Mosquera and P. Moreda, "SMILE: An informality classification tool for helping to assess quality and credibility in web 2.0 texts," in *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.
- [20] A. Gupta and P. Kumaraguru, "Credibility ranking of tweets during high impact events," in *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media (PSOSM)*. ACM, 2012.
- [21] F. Chevalier, S. Huot, and J.-D. Fekete, "Wikipediaviz: Conveying article quality for casual wikipedia readers," in *Pacific Visualization Symposium (PacificVis)*, 2010 IEEE. IEEE, 2010, pp. 49–56.
- [22] J. R. C. Nurse, S. Creese, M. Goldsmith, and K. Lamberts, "Trustworthy and effective communication of cybersecurity risks: A review," in *Workshop on Socio-Technical Aspects in Security and Trust (STAST) at The 5th International Conference on Network and System Security (NSS)*. IEEE, 2011.
- [23] D. C. Howell, *Fundamental statistics for the behavioral sciences*, 7th ed. Wadsworth Publishing Company, 2011.
- [24] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974.
- [25] P. Croskerry, "Achieving quality in clinical decision making: cognitive strategies and detection of bias," *Academic Emergency Medicine*, vol. 9, no. 11, pp. 1184–1204, 2002.
- [26] E. Peters, K. D. McCaul, M. Stefanek, and W. Nelson, "A heuristics approach to understanding cancer risk perception: contributions from judgment and decision-making research," *Annals of Behavioral Medicine*, vol. 31, no. 1, pp. 45–52, 2006.
- [27] A. Furnham and H. C. Boo, "A literature review of the anchoring effect," *The Journal of Socio-Economics*, vol. 40, no. 1, pp. 35 – 42, 2011.