Jurgen Willems and Carolin Waldner and John Ronquillo

Reputation Star Society: Are star ratings consulted as substitute or complementary information?

Article (Draft)

Original Citation:

Willems, Jurgen ORCID: https://orcid.org/0000-0002-4439-3948 and Waldner, Carolin and Ronquillo, John

(2019)

Reputation Star Society: Are star ratings consulted as substitute or complementary information?

Decision Support Systems, 124.

p. 113080. ISSN 0167-9236

Reputation Star Society:

Are star ratings consulted as substitute or complementary information?

Jurgen Willems – Vienna University of Economics and Bunsiness (WU)

Carolin J. Waldner – University of Hamburg

John C. Ronquillo – University of Colorado Denver

Abstract

To simplify decision making processes, online platforms frequently display reputation star ratings as an indication of the quality of a product, service, or organization. Can information provided by such star ratings draw away attention from other information? This is an important question for platform developers to adjust the use of such ratings. We conduct a between-subjects laboratory experiment (n = 121) where we manipulate the difference between the reputation star ratings of two social profit organizations, and ask respondents to indicate which organization they prefer. Applying eye-tracking technology, we analyze how the visual attention between the treatment conditions differs. Our findings show that reputation star ratings are consulted as complementary information, rather than as substitute information. Moreover, the results suggest that the lack of stars – not the presence of more stars – attracts visual attention.

Key words: Online reputation systems; star ratings; eye-tracking; decision-making

Declarations of interest: none

Preprint: this is a preprint version of the article: Willems, Jurgen, Carolin J.Waldner, and John C.Ronquillo (2019) Reputation Star Society: Are star ratings consulted as substitute or complementary information? Decision Support Systems. In Press, Corrected Proof. Available online 30 June 2019, 113080. https://doi.org/10.1016/j.dss.2019.113080

Acknowledgements:

This research was funded by the Deutsche Forschungsgemeinschaft (DFG-WI4542). For the overall overview of the project, we refer to: Willems, J., & Waldner, C. J. (2019). Reputation Management for Nonprofit Organizations. Open Science Foundation: https://osf.io/wj2ep ; DOI: 10.31219/osf.io/wj2ep

INTRODUCTION

Decision making based on reputation star ratings has become a substantial part of our personal and professional lives (Yang, Hu, & Zhang, 2007). When booking a hotel, buying a book, choosing a movie, or selecting a restaurant on one of many online platforms, information is often given in the form of reputation ratings that indicate how other users and/or external specialists have rated the product or service (Jøsang, Ismail, Audun, & Boyd, 2007; Zacharia, Moukas, & Maes, 2000). These reputation ratings are often presented visually, by colored stars or checkboxes, making them easy to identify and interpret. Such reputation ratings – which we refer to in this study as reputation star ratings – have gained great importance since the development of the Internet and especially the Web 2.0 and 3.0 revolutions. Despite the fact that reputation star ratings can provide information that may ease decisions in an overloaded information society, their derivation and underlying content is often is often — either intentionally or unintentionally — nontransparent (Zhang, 2006).

In this study, we start from the assumption that this reputation information can be interpreted as substitute or as complementary information compared to other available, non-reputation information. If reputation star information is perceived as substitute information, a disadvantage might be that differences in such ratings between two comparable objects mask other information or draw away attention from other information that could also be relevant for particular people and/or in particular contexts. Decision makers might neglect or ignore other information, particularly in a setting where reputation star ratings show a substantially clear difference between objects. Only when the reputation star ratings perfectly summarize the other information, if even possible, such substituting tendency could indeed make the information search and decision process more efficient.

In contrast, decision makers might interpret reputation star ratings as additional or complementary information to other available non-reputation related information. This means that differences in reputation star ratings are just a part of all available information, and in fact could potentially even trigger more attention to other relevant information and result in a more thorough decision making process. However, if reputation star information is meant to summarize the other available information, a consultation process, in which it is considered as complementary information, might take longer than necessary. Considering these contrasting perceptions of the role of reputation star ratings in comparison to other available information, it is relevant for decision system developers to know how such simple star ratings influence visual attention patterns to reputation star ratings, but also to other available information. The question is thus: Are reputation star ratings consulted as complementary or as substitute information by decision makers? Or in other words, does a difference in reputation star ratings draw away attention from non-reputation information? Answering this question can give insight into how reputation star ratings influence information consultation patterns. This can help decision system developers adjust their use of reputation star ratings, depending on their intended purpose with respect to providing reputation star ratings as a part of all available information.

We conducted a laboratory experiment involving eye-tracking analysis to answer this question. In a situation where decision makers compare two objects to indicate which they like most, we test how a clear difference in stars influences the overall visual attention pattern, compared to a situation where there is no difference in stars. Our design reflects a simplified setting of the more complex comparisons that often happen in daily life. We designed a two-group between-subjects experiment in which participants compare only two objects. We manipulated only the difference in stars between treatment groups, and kept constant all other potential factors that can influence visual attention. This allows us to test an important potential effect of reputation star ratings with more statistical power and internal validity (Orquin & Holmqvist, 2018). Our design gives an initial, preliminary answer to our research question, and it opens potential avenues for further research.

We identified eye-tracking analysis as a relevant method for our research question, given its value in analyzing visual attention on a granular level (Schall & Bergstrom, 2014; Shojaeizadeh, Djamasbi, Paffenroth, & Trapp, 2018). As our main variable of interest is visual attention to different pieces of information, eye-tracking analysis can provide relatively high-quality observational data from decision makers. For this research, it is important that eye-tracking is unobtrusive, accurate in capturing attention to visual stimuli, and provides an objective measurement of visual patterns, as opposed to potential subjectivity and exposure to other biases in other types of methods (Djamasbi, 2014; Djamasbi & Hall-Phillips, 2014; Liu, Djamasbi, Trapp, & Shojaeizadeh, 2018). A substantial body of literature has been developed on how this fine-grained visual attention to specific pieces of information is related to concrete decisions based on the information provided (for an overview see Schulte-Mecklenbeck, 2017). The main focus here, however, is not on the relationship between visual attention and the actual decision, but on how a particular type of information (reputation star ratings) influences the overall attention pattern, which is crucial in the overall decision process.

In this study we contribute first to the general understanding of how users consult star rating information, building on previous observational and experimental research on reputation, advice, and decision-making (Dalal & Bonaccio, 2010; Fuller, Serva, & Benamati, 2007; Kim, Ferrin, & Rao, 2008; Weigelt & Camerer, 1988; Yaniv & Kleinberger, 2000). This study takes a user-centered and behavioral approach to understand how heuristic information signals in online markets and platforms influence perceived quality and reputation (Fuller et al., 2007; Kim et al., 2008; Qi, Footer, Camerer, & Mobbs, 2018).

As a second main contribution, this study tests a specific guidance design feature in a common decision support system (Morana, Schacht, Scherp, & Maedche, 2017). Guidance design features are any technical or behavioral features that can help decision makers in their decision (for an extensive overview and taxonomy of guidance design features, we refer to Morana et al. 2017). From that perspective, we provide an example for other studies to continue to explore how different rating systems, as a particular form of a guidance design feature, can support decision making (Dalal & Bonaccio, 2010; Karimi & Wang, 2017; Morana et al. 2017).

Third, this study answers the call for the application of more innovative research methods and approaches to better understand decision processes that are guided by decision support systems, and reputation rating systems in particular (Fehrenbacher & Djamasbi, 2017; Tadelis, 2016).

BACKGROUND - REPUTATION STAR RATINGS AS COGNITIVE SIGNALS

Reputation star ratings often form the crucial control and trust mechanisms in secondary exchange or 'sharing economy' applications, such as AirBnB and Uber (Muchahari & Sinha, 2018). Similarly, such star ratings are also increasingly used (Banerjee, Bhattacharyya, & Bose, 2017; Fombrun, 2007) to assess the investment-worthiness of crowdfunded startup projects (e.g. Kickstarter, Indiegogo), the impact of charities or social profit organizations (e.g. Charity Navigator, Better Business Bureau Wise Giving Alliance), the attractiveness of universities (e.g. US News and World Report, Shanghai Ranking/Academic Ranking of World Universities), or the impact of scientists (e.g. Scopus and ResearchGate).

A substantial body of interdisciplinary literature has been developed on reputation star ratings and decisions (Chen, Zheng, Xu, Liu, & Wang, 2018; Zhou, Dresner, & Windle, 2008). Reputation star ratings function as important signals that can influence to varying degrees our cognition of the perceived value of several objects, such as products, services, organizations, and people (George, Dahlander, Graffin, & Sim, 2016). With this study, we build on literature that has focused on the relation between available information and how it is presented on the one hand, and a decision maker's decision on the other. Given the prevalence of information presented online, the process of discerning information available on websites is a cognitive one. This process can be influenced by design, text, and graphics through a visual hierarchy of information, with graphical representations and imagery used to draw attention to a specific area, in preference over text (Faraday, 2000; Grier, Kortum, & Miller, 2007; Djamasbi, Siegel, & Tullis, 2012).

These earlier studies have shown the strong relationship between visual attention to information and actual decisions. However, there is scant research that shows how a particular type of information might influence the visual attention pattern to all available information, which can affect the final decision. To improve our understanding of the potential influence of reputation star ratings on the distribution of visual attention to all available information, our hypotheses and analyses focus on the relationship between what a reputation star rating is reporting (i.e. different amount of stars versus equal amount of stars), and the visual attention to that information as well as to other, non-reputation information.

HYPOTHESES: DO DIFFERENCES IN STAR RATINGS DRAW AWAY ATTENTION FROM OTHER INFORMATION?

With respect to cognitive information search and process patterns, findings are often summarized in two types of processes: top-down and bottom-up control of attention (Orquin, Ashby, & Clarke, 2016; Orquin, Bagger, & Loose, 2013; Orquin & Loose, 2013; Theeuwes, 2010). Top-down control describes active attention patterns that are intentional and goal related, for example when a person is given a very particular task with respect to the information provided, or when motivation for the task is high (Orquin et al., 2013; Theeuwes, 2010). Bottom-up control is attention-grabbing or stimuli-driven, meaning that the formatting of the information (e.g. text size, color, or logo contrast with background) draws the initial passive attention of a person. Extensive empirical evidence suggests that initial attention to newly presented information is completely driven by bottom-up control, but over time (in as little as a matter of milliseconds) top-down control takes over (Theeuwes, 2010).

In general, we can assume that the presence of stars will attract bottom-up attention. Summarized by the theory of visual hierarchy (Faraday, 2000), information that stands out because of color, size, and/or graphical design, serves as an entry point in the information scanning process of decision makers (Djamasbi et al., 2012). This means that decision makers sort out visual key cues in the information field that is available to them, and implicitly attribute weights to these pieces of information for further evaluation in their overall decision (Djamasbi, 2014).

When a difference in stars exists, stakeholders are likely to notice and consider it immediately as relevant information for a decision. Therefore, a clear visual difference in stars is likely to also ignite a top-down control and enable the stakeholder to satisfy their top-down visual search pattern. While decision makers process the reputation information based on the difference in stars, they are likely to pay their visual attention to the reputation star rating (Theeuwes, 2010; Awh et al., 2012). As a result, differences in reputation stars will attract longer visual attention processing times based on the difference in stars. Therefore, our first hypothesis is:

> *Hypothesis 1: In a comparison of two objects including reputation star ratings, a difference in stars will result in more visual attention to the star rating information, compared to when there is no difference in reputation stars.*

Building on the rationale for Hypothesis 1, we then focus on what happens with visual attention to other types of information as a result of differences in stars. This allows us to better understand whether reputation star ratings draw away attention from other, non-reputation information. We rely on ecological rationality theory to explain how respondents attribute importance to reputation star ratings, and how a difference in stars could reduce perceived value of other available information – and thus less visual attention (Smith, 2003; Todd & Brighton, 2016). Ecological rationality theory states that humans draw on a broad set of decision mechanisms and trade off the expected effort and outcome, depending on their specific contextual goals and the importance of those goals (Gigerenzer, 2008; Smith, 2003). Such trade-offs help people deal efficiently with various types of information at once (Gigerenzer & Goldstein 1996). A crucial argument in ecological rationality theory is that conscious human beings continuously and recurrently evaluate the context in which they make decisions and the information available for the decision. Based on that evaluation, decision makers episodically determine how much cognitive effort they are willing to additionally attribute to reach a final decision.

For this study, the top-down element of visual attention to reputation star information is based on the assertion that a difference in stars, either partially or completely, provides the necessary information for decision makers to choose the best object. This means a difference in stars has certain informational value for the decision maker in the decision process. Therefore, and according to ecological rationality theory, when a difference in stars is observed, decision makers might attribute more importance to the reputation star rating and perceive the non-reputation information as less relevant and pay less visual attention to it.

From this perspective, reputation star ratings and other reputation-related information can also have a detrimental role in the decision maker's use of information by potentially drawing away attention from other non-reputation information. The following two reasons explain this trade-off. First, reputation often functions as a simplified summary of a broad range of potentially relevant elements in the decision making process (Dalal & Bonaccio, 2010; Fuller et al., 2007; Kim et al., 2008; Puncheva, 2008). This means decision makers see reputation-reporting metrics as a proxy for overall quality or performance of a product, service, or organization. Reliance on heuristic information signals like reputation star ratings is a rationally efficient approach when quality and performance are multi-dimensional, subjective, and opaque (Gigerenzer, 2008; Park & Nicolau, 2015).

Second, reputation is the socially constructed set of expectations that multiple stakeholders hold about an object or person and that they take into account for their decisions, such as buying a product, recommending a service, or donating to a non-profit (Michel & Rieunier, 2012; Mishina et al., 2012). Hence, the social dimension of reputation, or the fact that reputation is mainly related to what several others have experienced, provides decision makers with additional, indirect information that can justify a reduction of one's own mental efforts – for example by paying less attention to other available information – while still expecting to make a good decision (Gigerenzer & Brighton, 2009).

The reliance on reputation ratings and the reduced attention to non-reputation information might be particularly relevant in situations typified by asymmetric information (Puncheva, 2008; Weigelt & Camerer, 1988). In a highly marketized and digitized era where information is often sought online, individuals often forego in-person, physical interactions. For example, Weigelt and Camerer (1988) assert that consumers cannot distinguish between high- and low-quality goods before consumption, and therefore question whether producers are able to credibly convey quality. This, again, reinforces the importance of reputation information. While individuals can use reputation as a screening strategy, they do so with imperfect or incomplete information (Connelly et al., 2011; Kivetz & Simonson, 2000; Weigelt & Camerer, 1988). Absent perfect or complete information, the individual is likely to engage in some risk when choosing by placing her trust in the limited information available, or potentially conflicting information when information is available through multiple channels (e.g. reputation systems such as Yelp, Better Business Bureau, etc.) (Fuller et al., 2007; Kim et al., 2008). In lieu of direct experience, individuals are likely to screen available choices in order to make the best decision for them, forming trusting beliefs based on what they know from the reputation information (Fuller et al., 2007; Puncheva, 2008). Where ratings are similar, it is plausible that information seekers and decision makers will go beyond the reputation star ratings for more contextual information. Where ratings starkly differ, it is likely that the individual may accept the better or higher rated option to increase the likelihood of a better personal pay-off. In other words, the trade-off between expected effort and outcome causes the individual to place limited attention to contextual information, if the difference in reputation star ratings helps to make a decision.

Due to the specific nature of reputation and reputation-related metrics in combination with this ecological rationality perspective, we expect:

> *Hypothesis 2: In a comparison of two objects including reputation star ratings, a difference in stars will result in less visual attention to other available, non-reputation information (compared to a situation with no difference in stars).*

Our hypotheses focus on the fact that a difference in stars will lead to more visual attention to reputation star information, and to less visual attention to non-reputation information. The combined interpretation of the results for both hypotheses can inform us how visual attention switches between types of information. In the case of both hypotheses being supported – and regardless of whether a difference in stars changes the attention duration to any type of information – we can observe a change in the overall attention pattern as a result of a difference in stars, with less relative attention to other available information. We must note that we are interested in how the attention of decision makers changes among different types of information that are available and compared across the object, and thus, the hypotheses do not focus on the object that will be preferred as a result of a difference in star ratings, or which object receives more or less visual attention. However, we provide a complementary analysis for the latter to frame our study in the larger body of literature on visual attention.

EXPERIMENT

Experimental design

The design of this study is consistent with the ethical standards of the experiments laboratory of the Faculty of Business, Economics, and Social Sciences at the University of Hamburg (WISO laboratory). Before entering the laboratory, students were informed that the study involved eye-tracking analysis, a mechanism by which eye movements are detected to focus on specific areas on a computer screen where information may be displayed. All respondents were informed during the completion of their consent forms about the general purpose of the experiment and the voluntary and anonymous nature of their participation.

All respondents completed a maximum of three calibration rounds to adjust the eye-tracking system (Tobii Pro X2-60 eyetracker) for the information displayed on a 24-inch computer screen. This calibration contained an initial basic task where respondents were asked to focus on five dots that appear in sequence on different places on the screen. In doing so, the eye-tracking recording can be matched with what is seen on the screen, and before continuing the recording for the actual experiment, it can be determined whether the eye-tracking recording is functioning properly. This is a standard step to improve data quality in all eye-tracking experiments conducted in the WISO laboratory of the University of Hamburg. When the calibration did not work, we asked respondents to do another round. No more than three attempts were needed for any of the respondents. After the calibration, the participants started with a computer survey in which the experimental treatment was embedded. This survey was conducted with Qualtrics Surveys.

Our experiment focused explicitly on testing the hypotheses formulated, so that only a difference in reputation star ratings was altered between treatment groups in which participants were randomly assigned and all other available information and its location on the screen were kept constant. Mainly concerned about potential threats to (external) validity for eye tracking studies (Orquin & Holmqvist, 2018), we focused on this specific causal element, and with a sufficient sample size. As a result, our design does not allow for extensive exploration of other potential moderating variables.

Framing and decision

The experiment started with an introduction and decision task, which was as follows: “On the following page you will find information from two social profit organizations. Please choose the organization you would like to support the most (e.g. through social media endorsement, donations and/or volunteering).” On the next page, the participants then received information about the two organizations, accompanied with the specific direction: “Please select the organization you prefer.” The two organizational fact sheets were presented next to each other with basic information about (1) organizational mission, (2) social impact, expressed with a reputation star rating, (3) number of volunteers, and (4) annual budget. Organization A was always presented on the left, while Organization B was always presented on the right. Only the amount of stars was altered left or right on the screen. Figure A1 in the Appendix gives an overview of these fact sheets and how they were presented to the two treatment groups.

All respondents were randomly assigned to one of two groups (based on the embedded randomization mechanism in Qualtrics Surveys). In the first group (Treatment group 1; n = 61), both organizations had an equal star rating (4 out of 5 stars). In the second group (Treatment group 2; n = 60), there was a difference in stars (one organization had 1 star and the other had 5 stars, out of 5). The second group was further split up into a group where Organization A had more stars than Organization B (n = 30), and another group where Organization B had more stars (n = 30). This means that for the group with unequal stars, 50 percent of the respondents saw more stars on the left, and 50 percent saw more stars on the right. This was done because visual attention patterns often suffer from a reading bias, meaning that information on top and left are more frequently consulted (at least in societies with a language that is read from left to right) (Djamasbi, Siegel, & Tullis, 2011; Dallas, Liu, & Ubel, 2019). We controlled for this reading bias by design to ensure that in a situation with unequal stars (Hypothesis 1), a difference in visual attention can be attributed to the mechanisms explained in our theory section and not to the specific location of information on the screen (e.g. left versus right). With respect to other design features of our information sheet, we also refer to them in the next subsection where we discuss visual attention as our dependent variable.

Respondents could choose their preferred organization by clicking the tick box underneath the respective organizational fact sheet. Respondents could consult all available information for an unlimited amount of time before making a decision and moving to the next survey page (Orquin & Holmqvist, 2018).

Both organizations were fictitious to avoid biases based on prior knowledge and experience of a known organization that might have influenced respondents' visual attention and final decision. We chose the particular setting for the following three reasons. First, social profit organizations are characterized by the fact that their performance is often (1) very subjective, i.e. highly dependent on stakeholder perceptions, (2) is multi-dimensional, (3) hard to quantify, and (4) information about organizational performance is often highly asymmetric (Lecy, Schmitz, & Swedlund, 2012; Willems, Boenigk, & Jegers, 2014). This means that the organization's effectiveness reputation, i.e. the social construction of what people think their social value is for society, often plays a major role in how stakeholders make decisions about such organizations. Given the theoretical focus on the socio-cognitive nature of reputation makes this setting highly interesting, particularly for Hypothesis 2.

Second, we believe this setting has high practical relevance, as many social profit organizations struggle in communicating performance and effectiveness to broad audiences, especially as they are increasingly pushed toward reporting more quantifiable outputs (Willems, Jegers, & Faulk, 2016). Answering our research question in this particular context can inform nonprofit managers and policy makers whether the evolution towards reporting social impact performance based more on single metrics is a good evolution, as it might reduce stakeholder attention to other relevant information, that, in particular, might be relevant in a context where multi-dimensional performance is a core element.

Third, as we rely on a student sample (see the following subsection), the question on which social profit organization they like most is a realistic question they can answer that is neither too hypothetical nor in an unrealistic setting. This is why we provided additional clarification on why we asked their preference and added the context of potentially supporting the organization. In this context, we rely on the extensive body of theoretical and empirical literature that shows how positive attitudes are a prerequisite for actual supporting behavior (Bendapudi, Singh, & Bendapudi, 1996; Babin & Bruns, 1997). Nevertheless, for our design and the hypotheses we test, we want to clarify that the focus is not on testing the effect of our treatment on actual behavior, but that the goal was to give the participants a decision task in which they consult all the available information.

Participants

Students from the University of Hamburg were invited to participate in the laboratory experiment in which they were paid 10 euro-per-hour. We aimed for about 60 students per treatment group (in our main two-group comparison). This means we were able to discover a medium effect size, in case a true effect exists (based on the following a priori assumptions: sample size calculation for a two group ANOVA test, Cohen's $f = 0.25$, i.e. Cohen's d = 0.50 medium effect size, $\beta = 0.80$ and $\alpha = 0.05$) (Champely, 2018). In total, 129 respondents participated, and after data quality checks, 121 responses could be used for this experiment (74 identified as female, and the average age was 25.87, s.d. = 4.62). In the other eight cases, some eye-tracking data were missing (for example, participants wore glasses or looked away from the computer screen for too long during the experiment) or were not properly calibrated with questionnaire data.

Variables

The focus of the analysis is on visual attention to the information in the organizational fact sheets, with special attention on two variables: (1) visual attention to reputation star rating, and (2) visual attention to the other, non-reputation information. Visual attention is operationalized as the total fixation time (in seconds) that respondents focus on particular areas of interest (AOIs) in the organizational fact sheets. We applied the Tobii Studio default settings: average eye selection, I-VT filter (velocity threshold of 30 degrees per second), merged adjacent fixations of a maximum time between fixations of 75 milliseconds, and maximum angle between fixations 0.5 degrees; and a minimum fixation duration of 60 milliseconds. Other studies suggest longer minimum fixation durations, while Orquin and Holmqvist (2018) suggest not too long of a minimum fixation duration to have sufficient data quality.

Orquin, Ashby and Clarke (2016) point out that no standardized procedures exist to design information sheets or mark AOIs within them. Using their suggestions, the organizational fact sheets were designed in such a way to ensure that different information pieces had sufficient white space between them to avoid overlapping fixation distributions and thus limit the risk of false positive fixation registrations. Four AOIs were designed and placed on each organizational fact sheet. An overview of the AOIs and how they were combined, are presented in Figure A2 in the Appendix. The sum of visual attention to reputation star information constitutes the first main variable, and the sum of visual attention to the all other information constitutes the second main variable. Table A1 in the Appendix reports both variables for the overall sample and per treatment group. Respondents consulted both organizational fact sheets on average for 9.146 seconds (s.d. = 5.013), with a minimum of 0.92 seconds, and a maximum of 26.75 seconds. On average, respondents looked 4.683 seconds (s.d. = 2.862) at the reputation star info and 4.464 seconds (s.d. = 3.409) at the non-reputation information.

Furthermore, we examined and reported only on visual attention, measured as total fixation time. While various eye-tracking software applications provide a multitude of other eye-movement metrics, we focused on a single metric most suited for our particular research question. This is based on the methodological recommendation from Orquin and Holmqvist (2018) to avoid the analysis of too many eye-tracking metrics in a single study, as it is a threat to statistical validity (mainly multiple comparisons problem, especially due to restricted statistical power in limited sample sizes for eye-tracking studies).

Considering the other potential validity threats of eye tracking studies (Orquin & Holmqvist, 2018), we designed our experiment and conducted our analysis in a way that does not directly compare visual attention to different types of information (e.g. a direct comparison of written information versus pictured information). Instead, we test how the visual attention distribution between reputation star ratings information and non-reputation information changes, as a result of a difference in stars (between-subjects design).

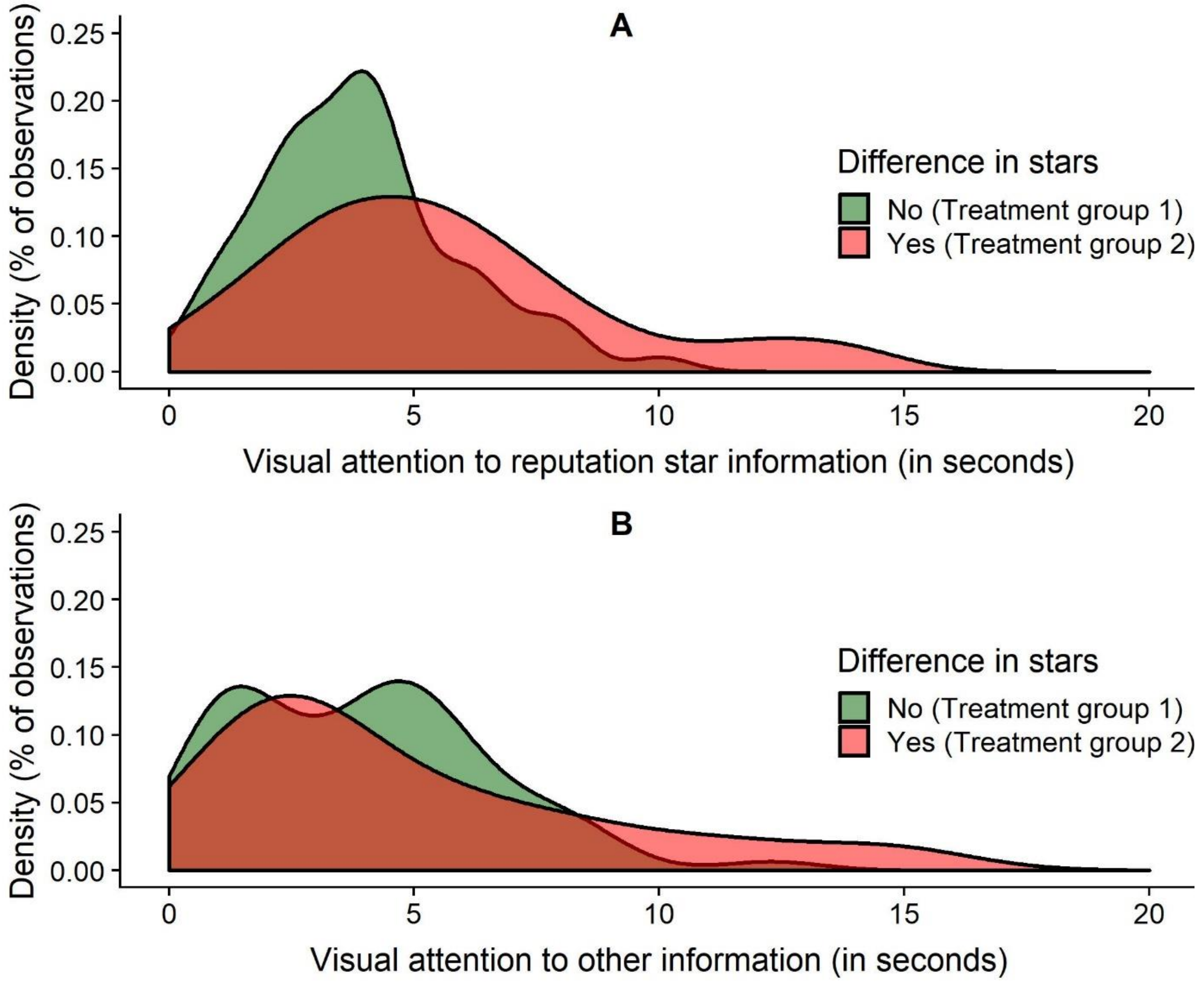
We test how this distribution changes between types of information across the objects that are compared by the decision makers. We start with the comparison of the total visual attention to reputation star information of both organizations (sum of total fixation time in areas AOI 3 and AOI 4, in Figure A2 in the Appendix) with the total visual attention to the other areas (AOI 1, AOI 2, AOI 5, AOI 6, AOI 7, and AOI 8, in Figure A2 in the Appendix, also summarized). This means that AOI 1, AOI 2, AOI 5, AOI 6, AOI 7, and AOI 8 were kept constant for all respondents of both treatment groups, and the absolute visual attention to these areas captures all the potential effects that might explain differences in visual attention for specific AOIs (explained in more detail in Orquin & Holmqvist, 2018, p. 1650), while a relative difference between treatment groups is a result of the single experimental manipulation in this study.

Results

Testing the hypotheses: visual attention

To understand whether reputation star ratings are consulted as complementary or substitute information, we analyzed the differences between visual attention placed to the respective AOIs. We consider it a complementary consultation if there is no significant difference in fixation time of the respective AOIs, and a substitute consultation, if respondents placed significantly less visual attention to the non-reputation information.

Figure 1: Visual attention in seconds to reputation star information and other information.



Note: Density plots (Wickham, 2016) report the relative frequencies (in percentages; vertical axis) of observations for visual attention (in seconds; Horizontal axis) to the reputation star information (part A) and to the non-reputation information (part B). The line covering the green area gives the relative frequencies of observations for absolute visual attention that a respondent in Treatment group 1 spent on a type of information (No difference in stars). The line covering the red area, gives the relative frequencies of observations for absolute visual attention (in seconds) that a respondent in Treatment group 2 spent on a type of information. For part A, the difference in visual attention to reputation star information between treatment groups is significant (Hypothesis 1: Kruskal-Wallis $\chi 2 = 7.6923$, df = 1, $p = 0.0055$). For part B, the difference in visual attention to other available, non-reputation star information between treatment groups is not significant (Hypothesis 2: Kruskal-Wallis $\chi 2 = 0.83723$, df = 1, $p = 0.3602$).

Figure 1 reports with density plots (Wickham, 2016) the relative frequencies of observations for visual attention (in seconds) to the reputation star information (Figure 1A) and to the non-reputation information (Figure 1B). The lines give the relative frequencies of observations (in percentages) for absolute visual attention (in seconds) that a respondent spent on a piece of information. For Figure 1A, this is visual attention to reputation star ratings; For Figure 1B, this is visual attention to the other available, non-reputation information. We test the difference in averages between the two treatment groups with a Kruskal-Wallis rank sum test, as it is most appropriate for a multi-group comparison of a variable with a non-normal distribution (see Figure 1). The group of respondents with a difference in reputation stars (red area, in Figure 1) on average spent significantly more attention on the reputation star rating (Table A1 in the Appendix: Group 2 mean = 5.49) than the group of respondents with equal reputation stars (Table A1 in the Appendix: Group 1 mean = 3.88) (Kruskal-Wallis $\chi^2 = 7.6923$, df = 1, p = 0.0055). As visible in Figure 1A, on average, respondents consulted the reputation star ratings longer when they reported an actual difference, i.e. the flatter but longer shape of the density curve covering the red area for Treatment group 2, compared to the density curve covering the green area for the respondents in Treatment group 1. Hence, Hypothesis 1 is supported.

A difference in reputation stars, however, did not result in more or less visual attention to the non-reputation information (Kruskal-Wallis χ^2 = 0.83723, df = 1, p = 0.3602; Group 1 mean = 3.94, Group 2 mean = 5.00). This is evident in Figure 1B, as the density curves and areas are mainly overlapping for both treatment groups. In other words, the distributions of visual attention to the other available information are not significantly different as a result of the experimental treatment. This suggests that a difference in reputation stars is consulted as additional, complementary information, rather than as substitute information. Thus, Hypothesis 2 is not supported.

Complementary analysis 1: a further exploration of visual attention to the subgroups in Treatment group 2

We can better understand the change in visual attention patterns as a result of our main treatment by further exploring differences between Treatment group 1, and the two subgroups of Treatment group 2. We report in Figures 4, for Treatment group 1 (equal amount of stars) and the two subgroups of Treatment group 2 (unequal amount of stars) group means and distributions for visual attention to different types of information. We use bean plots (Kampstra, 2008) or pirate plots (Phillips, 2017) to provide this additional information. Such plots give good insights in group means, but also in the (difference between) distributions of each group. The black horizontal line in each group reports the group mean (visual attention to an AOI or a combination of AOIs). The box around the mean line reports the 95% confidence intervals, and the (colored) density curves/clouds give insights into the distribution of observations within each group. This is relevant, as the data is not normally distributed. The wider the density cloud in a group, the more observations occurred for the corresponding value on the vertical axis. As a result, these plots not only make it clear how group means differ as a result of different experimental treatment groups and sub-groups, but also how internal distributions might differ.

Figure A3 in the Appendix, Part I reports visual attention to all reputation information (for both organizations combined), based on a three-fold division: (1) Organization A has more stars, (2) both organizations have equal stars, and (3) Organization B has more stars. Both sub-groups of Treatment 2, i.e. a different number of stars, have a significant higher mean than Treatment group 1. This is an alternative way of presenting the results for Hypothesis 1. For the three-group comparison, the Kruskal-Wallis chi-squared test shows a significant difference between groups (χ^2 = 8.7928, df = 2, p = 0.012; Group mean 'A and B have equal stars' = 3.88, Group mean 'A has more stars' = 5.75, Group mean 'B has more stars' = 5.24). However, this additional test and graph suggests that, for example, a reading bias, where decision makers spend more visual attention on information on the left side of an information sheet (Djamasbi et al., 2011), did not interfere with presenting more stars on the left compared to presenting them on the right. This means that more visual attention to reputation star ratings as a result of a difference in stars (Hypothesis 1) is not influenced by their specific location on the screen, at least from a left-versus-right perspective, keeping the top-versus-bottom element constant across all treatment groups and subgroups.

Figure A3 in the Appendix, Part II reports for the same three groups the visual attention to non-reputation information. This is an alternative presentation of the results for Hypothesis 2, but with a distinction for the subgroups of Treatment group 2. No significant difference is observed among the three groups (Kruskal-Wallis χ^2 = 0.88359, df = 2, p-value = 0.6429; Group mean 'A and B have equal stars' = 3.94, Group mean 'A has more stars' = 5.25, Group mean 'B has more stars' = 4.75), indicating that there is no substituting effect of a difference in reputation star rating on other types of information (recall Hypothesis 2 is not supported). The higher means for both subgroups of Treatment group 2 suggest that there might be a tendency that a difference in stars triggers extra attention to the other available information, rather than drawing away attention from it. Before arriving at this conclusion, however, an alternative theoretical logic and more extensive testing is needed.

Complementary analysis 2: Choosing the preferred organization and the relatedness with visual attention

Figure A3 in the Appendix, Parts III and IV report, also with bean plots, the differences in visual attention to reputation star ratings, for Organization A and Organization B, respectively. By making the additional distinction per organization, we can explore in more detail how a difference in stars results in more or less attention to the actual choices. No significant differences are observed for attention to reputation information of Organization A (Figure A3 in the Appendix, Part III: Kruskal-Wallis χ^2 = 2.114, df = 2, p-value = 0.3475; Group mean 'A and B have equal stars' = 2.31, Group mean 'A has more stars' = 2.65, Group mean 'B has more stars' = 3.04). However, a long density plot for the 'Organization B has more stars' group suggests that a tendency might exist for few stars attracting visual attention, rather than many stars. In other words, when Organization B has more stars than Organization A (and A fewer than B), people focus slightly more on the reputation star rating of Organization A. This effect is significant in the alternative combination (Figure A3 in the Appendix, Part IV) as visual attention to the star rating of Organization B is higher when Organization A has more stars, while Organization B has fewer (Kruskal-Wallis chi-squared = 11.068, df = 2, p-value = 0.00395; Group mean 'A and B have equal stars' = 1.57, Group mean 'A has more stars' = 3.09, Group mean 'B has more stars' = 2.21). These results suggest that a lack of stars, rather than many stars attracts visual attention. We acknowledge that this preliminary finding is the result of a post-hoc exploration in a sample of limited size. It can, however, inform us on potential further directions to understand how a difference in stars influences overall attention patterns to all available information, and thus, more extensive testing is needed in the future.

To further explore the relationship between visual attention to reputation star ratings and non-reputation information, we report in Table A2 in the Appendix a set of four logistic generalized linear models with the choice of a preferred organization as the dependent variable (Organization A is preferred (value = 0) or Organization B (value = 1)). In these models, the choice for a preferred organization is explained, based on the experimental treatment and visual attention to different pieces of information. The comparison of these models can illuminate how visual attention to particular information potentially mediates the relationship between the experimental treatment (a difference in stars) and the final decision (which organization is preferred). It gives insight into the specific features of the design and setting we used, which is relevant to evaluate the generalizability of these findings. Again, this part of our analysis is complementary; previous studies had more appropriate designs to test this particular relationship. Visual attention on one hand and the final decision on the other do not necessarily relate unidirectionally to each other as cause and effect. A substantial body of literature suggests that an initial decision is made during the information consultation process, and that further information consultation is a function of this preliminary first choice. For example, Fiedler & Glöckner (2012; 2015) argue that a preliminary decision is made at a certain point (even early on) and that later visual attention is the result of cognitive processes to further verify and support the decision. Thus, the analysis in Table A2 in the Appendix reports the correlation of the decision with visual attention, rather than a causal effect.

Model 1 in Table A2 in the Appendix shows an overall significant tendency to prefer Organization B over A (0.454, $p < 0.05$), based on all the available information and when we do not specify (1) whether a difference in stars was presented and (2) which organization then had more or less stars than the other organization. This can be seen as a benchmark with which to compare the other models. Model 2 tests whether more or less stars influences the final choice that respondents make on their preferred organization, not taking visual attention into account. The non-significant coefficients show that only the treatment in itself does not explain the choice for one of the two organizations. This means that the difference in stars is probably not the main decision criterion for most respondents and that a combination of all the available information is consulted to make a final decision. When focusing on visual attention to explain whether Organization B is preferred over Organization A (Models 3 and 4), there is a relationship between visual attention for the respective organizations and the choice made. More visual attention to Organization A is related to a lower likelihood to prefer Organization B (-0.503; $p < 0.01$ for attention to reputation information, and -0.272; $p < 0.05$ for attention to non-reputation information). More visual attention to Organization B, at least for non-reputation information, is related to a higher likeliness to choose Organization B (0.587; $p < 0.01$). While this analysis does not provide conclusive evidence for it, these findings are consistent with the idea that extra visual attention is given to the decision option that is the final choice (Fiedler & Glöckner, 2012; 2015). Therefore, when we combine the results of our main analysis and hypothesis testing with these complementary results, we assume that a difference in stars creates a partial change in the overall attention pattern, and that this pattern is the basis for the final decision of decision makers.

CONCLUDING DISCUSSION

This study explored a primary research question of how a difference in reputation star ratings influences visual attention to the reputation star ratings, but also to non-reputation information. The utility of user experiences, based on either positive or negative interactions, and reported through reputation star ratings, has become increasingly important in the decision-making processes of individuals (e.g., Banerjee et al., 2017; Jøsang et al., 2007). When individuals are asked to compare information between two choices, we observed through an experimentally manipulated difference in reputation star ratings that visual attention patterns are influenced. While previous experimental research has elucidated the relationship between information and decision making, this study has contributed more granular knowledge to how information may be utilized, and what variation in ratings attracts eye movement or visual attention (supporting for example Fehrenbacher & Djamasbi, 2017). We investigated preferences towards social profit organizations, as they often depend on positive reputations to attract resources, yet stakeholders face high information asymmetries to evaluate these organizations. As stated in the introduction, star ratings are a common way to demonstrate reputation despite limited information. Future research can and should investigate whether our findings are transferable to other contexts.

This study reveals that reputation star ratings draw more visual attention to the stars when they report a clear difference, supporting Hypothesis 1. This shows that a difference in star ratings provides relevant information for decision makers that extends their visual attention to the stars, indicating that they process the star ratings not only in an initial bottom-up process, but also in a top-down process (e.g., Theeuwes, 2010). The study also shows that reputation star ratings are not drawing away visual attention from non-reputation information and stakeholders of social profit organizations do not seem to perceive a difference in star ratings as a substitute of the non-reputation information. These findings contrast our initial expectation (Hypothesis 2) and indicate that despite the trade-offs in cognitive efforts that stakeholders make according to ecological rationality theory (e.g., Todd & Brighton, 2016), they do not evaluate the organization based on either reputation or non-reputation information, but assess all information available.

In contrast, our complementary analyses identify the need for further testing, verification, and elaboration of some new propositions. In fact, our results suggest (though not at a 5-percent significance level) that a difference in stars for a reputation star rating might also trigger more visual attention to the other available information (e.g. Figure A3 in the Appendix, Part IV). The difference in stars potentially works as a visual entry point (Faraday, 2000; Djamasbi, et al., 2012; Djamasbi, 2014), triggering curiosity that leads to a more thorough information search. This proposition, however, needs to be more robustly tested in future research to have the potential to make stronger conclusions.

Additionally, the results of our complementary analysis suggest that individuals are likelier to focus on a lack of stars, or poor reputation reviews, rather than higher reputation star ratings. The caveat, as was presented in our analysis, is that this attention does not signify a concrete cause and effect with regard to decision making. Our results are only indicative of where individuals place their attention, rather than any discrete predictive measures with regard to personal choice. The findings, however, are consistent with the idea that the choice that garners more individual attention is likelier the final choice of the individual.

Limitations

While our complementary analyses are a good basis to formulate further research avenues, identifying the limitations of this study can help set forth new directions for further verification and falsification. First, we have only hypothesized and tested a particular effect of a difference in stars, which is mainly visual attention to all available information. Our choice for a robust analysis of two related hypotheses enables us to say with more certainty something about the relation between reputation star ratings and visual attention, but we remained rather exploratory on how that, in turn, influences the actual final decision. We reported this in complementary analyses, but given our setting and design, it remains difficult to make assertions on the causal relation between visual attention and the final decision between objects. Other attention measures, along with complementary methods such as discrete choice experiments, interviews in which choices are explained in retrospect, or even longitudinal diary studies can potentially provide richer information to complement the preliminary findings and interpretations of this study. For example, the order of fixations and comparison processes over time (e.g. in Djamasbi et al., 2012; Djamasbi, 2014) would require substantially larger sample sizes, but could also substantially contribute to a better understanding of the causal and subsequent steps in how a difference in reputation stars influences visual attention.

Second, we chose a particular and hypothetical setting where we asked students to indicate which social profit organization they most prefer. We believe such a setting is highly relevant given the crucial role of reputation for social profit organizations and because such a question is realistic for a student sample. We acknowledge that this is a particular setting and specific decision task, and that caution is warranted when generalizations are made to other contexts, such as buying products or services.

Third, we acknowledge that many aspects might influence visual attention, such as the location of information, the way it is presented, and what concrete facts the information reports. We kept factors constant where we did not formulate formal hypotheses. We subdivided one of the treatment groups in order to balance the visualization of stars on the left and right sides to avoid reader bias that could have interfered with the testing of our hypotheses. However, reputation star ratings themselves can come in many forms and shapes, and so can all other types of information. Its location and graphical design can also influence visual attention patterns. As such, further research should continue to test the hypotheses with other types of reputation stars, in other orders, and with a variety of other available information. In doing so, important moderating and/or confounding variables can be identified and tested in a way that potentially mitigate the hypothesized effects. This is particularly relevant to further exploration from a practical point of view, as it not only can give valuable insight on the value of reputation star ratings, but also on how they can be best designed and for what purposes. Eventually, this could lead to a more holistic framework or a set of practical guidelines that give a more complete insight in how different types of reputation star ratings can be best used by decision system developers. Potentially, different types of star ratings should be used depending on a broad range of contextual factors, such as the nature of other information available, the purpose of the types of information that is provided, and the type of decision to be made.

Finally, we focus only on a comparison of two objects in a controlled environment. As previously noted, this helped in improving the internal validity of our study (Orquin & Holmqvist, 2018), but given the multitude of options that are often presented on online platforms, this two-object choice lacks external validity. Therefore, further research should seek to find a good balance between internal and external validity when trading off highly realistic settings on one hand (Rayner, 2009) with sufficient statistical robustness on the other (Orquin & Holmqvist, 2018).

Practical implications

The main implication of our study is for decision support developers. Our preliminary results suggest that reputation star ratings are not considered as substitute information for other available information, even in a context where reputation is a crucial element of evaluation objects (i.e. social profit organizations). This means that reputation star ratings are consulted as additional information, and potentially trigger more attention to all information, including additional, non-reputation information. Decision system developers could therefore make sure that they do not rely just on positive star ratings but provide detailed information about the object, so stakeholders can draw on a variety of information to evaluate and make a decision, particularly when the amount of stars differs from other (competitive) objects.

REFERENCES

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus Bottom-up Attentional Control: A Failed Theoretical Dichotomy. *Trends in Cognitive Sciences, 16*(8), 437–43. https://doi.org/10.1016/j.tics.2012.06.010.

Babin, L. A., & Bruns, A. C. (1997). Effects of print ad pictures and copy containing instructions to imagine on mental imagery that mediates attitudes. *Journal of Advertising, 26*(3), 33–44.

Banerjee, S., Bhattacharyya, S., & Bose, I. (2017). Whose online reviews to trust? Understanding reviewer trustworthiness and its impact on business. *Decision Support Systems, 96*, 17–26. https://doi.org/10.1016/j.dss.2017.01.006.

Beldad, A., Snip B., & van Hoof, J. (2014). Generosity the second time around: Determinants of individuals' repeat donation intention. *Nonprofit and Voluntary Sector Quarterly, 43*(1), 144–163. https://doi.org/10.1177/0899764012457466.

Bendapudi, N., Singh, S. N., & Bendapudi, V. (1996). Enhancing helping behavior: an integrative framework for promotion planning. *Journal of Marketing, 60*(3), 33–49.

Champely, S. (2018). pwr: Basic functions for power analysis. R package version 1.2-2. Retrieved from https://cran.r-project.org/package=pwr.

Chen, R., Zheng, Y., Xu, W., Liu, M., & Wang, J. (2018). Secondhand seller reputation in online markets: A text analytics framework. *Decision Support Systems, 108*, 96–106. https://doi.org/10.1016/j.dss.2018.02.008.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management, 37*(1), 39–67.

Dalal, R. S., & Bonaccio, S. (2010). What types of advice do decision-makers prefer? *Organizational Behavior and Human Decision Processes, 112*(1), 11–23. https://doi.org/10.1016/j.obhdp.2009.11.007.

Dallas, S. K., Liu, P. J., & Ubel, P. A. (2019). Don't Count Calorie Labeling Out: Calorie Counts on the Left Side of Menu Items Lead to Lower Calorie Food Choices. *Journal of Consumer Psychology, 29*(1), 60–69.

Djamasbi S., Siegel M., Tullis T. (2011). Visual Hierarchy and Viewing Behavior: An Eye Tracking Study. In Jacko, J.A. (Ed.) *Human-Computer Interaction. Design and Development Approaches*. HCI 2011. Lecture Notes in Computer Science, 6761. Springer, Berlin, Heidelberg.

Djamasbi, S. (2014). Eye Tracking and Web Experience. *AIS Transactions on Human-Computer Interaction, 6*(2), 37–54.

Djamasbi, S., & Hall-Phillips, A. (2014). Visual search. In Schall, A., & Bergstrom, J. R. (Eds.*), Eye Tracking in User Experience Design*, pp. 27–43. Waltham, MA: Morgan Kaufmann.

Djamasbi, S., Siegel, M., & Tullis, T. (2012). Designing noticeable bricklets by tracking users' eye movements. In 45th *Hawaii International Conference on System Sciences*, 525–532. IEEE. https://doi.org/10.1109/HICSS.2012.200.

Faraday, P. (2000). Visually Critiquing Web Pages. *Proceedings of the 6th Conference on Human Factors and the Web,* 1–13. Austin, Texas.

Fehrenbacher, D. D., & Djamasbi, S. (2017). Information systems and task demand: An exploratory pupillometry study of computerized decision making. *Decision Support Systems*, *97*, 1–11. https://doi.org/10.1016/j.dss.2017.02.007.

Fiedler, S., & Glöckner, A. (2012). The dynamics of decision making in risky choice: An eye-tracking analysis. *Frontiers in Psychology, 3*(335).

Fiedler, S., & Glöckner, A. (2015). Attention and moral behavior. *Current Opinion in Psychology, 6,* 139–144.

Fombrun, C. J. (2007). List of lists: A Compilation of international corporate reputation ratings. *Corporate Reputation Review, 10*(2), 144–153. https://doi.org/10.1057/palgrave.crr.1550047.

Fuller, M. A., Serva, M. A., & Benamati, J. S. (2007). Seeing is believing: The transitory influence of reputation information on e-commerce trust and decision making. *Decision Sciences, 38*(4), 675–699. https://doi.org/10.1111/j.1540-5915.2007.00174.x.

George, G., Dahlander, L., Graffin, S. D., & Sim, S. (2016). Reputation and status: Expanding the role of social evaluations in management research. *Academy of Management Journal, 59*(1), 1–13.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science, 3,* 20–29.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1*, 107–143.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 640–669.

Grier, R., Kortum, P., & Miller, J. (2007). How Users View Web Pages: An Exploration of Cognitive and Perceptual Mechanisms. In Zaphiris, P. & Kurniawan, S. (Eds.) *Human Computer Interaction Research in Web Design and Evaluation*, 22–41. Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-59904-246-6.ch002.

Jøsang, A., Ismail, R., Audun, & Boyd, C. (2007). A survey of trust and reputation systems for online service provision. *Decision Support Systems, 43*(2), 618–644. https://doi.org/10.1016/j.dss.2005.05.019

Kampstra, P. (2008). Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. *Journal of Statistical Software, 28*(1), 1–9. https://doi.org/10.18637/jss.v028.c01.

Karimi, S., & Wang, F. (2017). Online review helpfulness: Impact of reviewer profile image. *Decision Support Systems, 96*, 39–48.

Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support Systems*, *44*(2), 544–564. https://doi.org/10.1016/j.dss.2007.07.001.

Kivetz, R., & Simonson, I. (2000). The effect of incomplete information on consumer choice. *Journal of Marketing Research, 37*(4), 427–448.

Lecy, J. D., Schmitz, H. P., & Swedlund, H. (2012). Non-governmental and not-for-profit organizational effectiveness: A modern synthesis. *Voluntas: International Journal of Voluntary and Nonprofit Organizations, 23*(2): 434–457.

Liu, W., Djamasbi, S., Trapp, A. C., & Shojaeizadeh, M. (2018). Measuring Focused Attention Using Fixation Inner-Density. In Schmorrow, D. & Fidopiastis, C. (Eds) *Augmented Cognition: Users and Contexts*. AC 2018. Lecture Notes in Computer Science, 10916, 105–116. Springer, Cham.

Michel, G., & Rieunier, S. (2012). Nonprofit brand image and typicality influences on charitable giving. *Journal of Business Research, 65*(5), 701–707. https://doi.org/10.1016/j.jbusres.2011.04.002.

Mishina, Y., Block, E. S., & Mannor, M. J. (2012). The path dependence of organizational reputation: How social judgment influences assessments of capability and character. *Strategic Management Journal*, *33*(5), 459–477. https://doi.org/10.1002/smj.958.

Morana, S., Schacht, S., Scherp, A., & Maedche, A. (2017). A review of the nature and effects of guidance design features. *Decision Support Systems*, 97, 31–42. https://doi.org/10.1016/j.dss.2017.03.003.

Muchahari, M. K., & Sinha, S. K. (2018). Reputation-based trust for selection of trustworthy cloud service providers. *Proceedings of the International Conference on Computing and Communication Systems, 24*.

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior research methods, 50*(4), 1645–1656.

Orquin, J. L., & Loose, S. M. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica, 144*(1), 190–206.

Orquin, J. L., Ashby, N. J. S., & Clarke, A. D. F. (2016). Areas of interest as a signal detection problem in behavioral eye-tracking research. *Journal of Behavioral Decision Making, 29*(2–3), 103–115.

Orquin, J. L., Bagger, M. P., & Loose, S. M. (2013). Learning affects top down and bottom up modulation of eye movements in decision making. *Judgment and Decision Making, 8*(6), 700–716.

Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research, 50*(67), 67–83.

Phillips, N. (2017). yarrr: A Companion to the e-Book "YaRrr!: The Pirate's Guide to R." Retrieved from https://cran.r-project.org/package=yarrr.

Puncheva, P. (2008). The role of corporate reputation in the stakeholder decision-making process. *Theory and Decision*, *72*(2), 221–243. https://doi.org/10.1007/s11238-011-9245-8.

Qi, S., Footer, O., Camerer, C. F., & Mobbs, D. (2018). A collaborator's reputation can bias decisions and anxiety under Uncertainty. *Journal of Neuroscience, 38*(9), 2262–2269. https://doi.org/10.1523/JNEUROSCI.2337-17.2018.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506.

Schall, A., & Bergstrom, J. R. (2014). Introduction to Eye Tracking. In Bergstrom, J. R., & Schall, A. (Eds) *Eye Tracking in User Experience Design,* pp. 3-26. Waltham, MA: Morgan Kaufmann.

Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., & Willemsen, M. C. (2017). Process-Tracing Methods in Decision Making: On Growing Up in the 70s. *Current Directions in Psychological Science*, *26*(5), 442–450.

Shojaeizadeh, M., Djamasbi, S., Paffenroth, R. C., & Trapp, A. C. (2019). Detecting task demand via an eye tracking machine learning system. *Decision Support Systems, 116*, 91–101.

Smith, V. L. (2003). Constructivist and ecological rationality in economics. *American Economic Review, 93*, 465–508.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics, 8*, 321–340. https://doi.org/10.1146/annurev-economics-080315-015325.

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta Psychologica, 135*(2), 77–99.

Todd, P. M. & Brighton, H. (2016). Building the theory of ecological rationality. *Minds and Machines, 26*, 9–30.

Weigelt, K., & Camerer, C. (1988). Reputation and corporate strategy: A review of recent theory and applications. Strategic Management Journal, 9(5), 443–454. https://doi.org/10.1002/smj.4250090505.

Wickham. H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Willems, J., Boenigk, S., & Jegers, M. (2014). Seven trade-offs in measuring nonprofit performance and effectiveness. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, *25*(6), 1648–1670.

Willems, J., Jegers, M., & Faulk, L. (2016). Organizational effectiveness reputation in the nonprofit sector. *Public Performance and Management Review. 39*(2), 476–497.

Willems, J., & Waldner, C. J. (2019). Reputation Management for Nonprofit Organizations. Open Science Framework : https://osf.io/wj2ep ; DOI: 10.31219/osf.io/wj2ep

Willems, Jurgen, Carolin J.Waldner, and John C.Ronquillo (2019) Reputation Star Society: Are star ratings consulted as substitute or complementary information? Decision Support Systems. In Press, Corrected Proof. Available online 30 June 2019, 113080. https://doi.org/10.1016/j.dss.2019.113080

Yang, J., Hu, X., & Zhang, H. (2007). Effects of a reputation feedback system on an online consumer-to-consumer auction market. *Decision Support Systems*, *44*(1), 93–105. https://doi.org/10.1016/j.dss.2007.03.005.

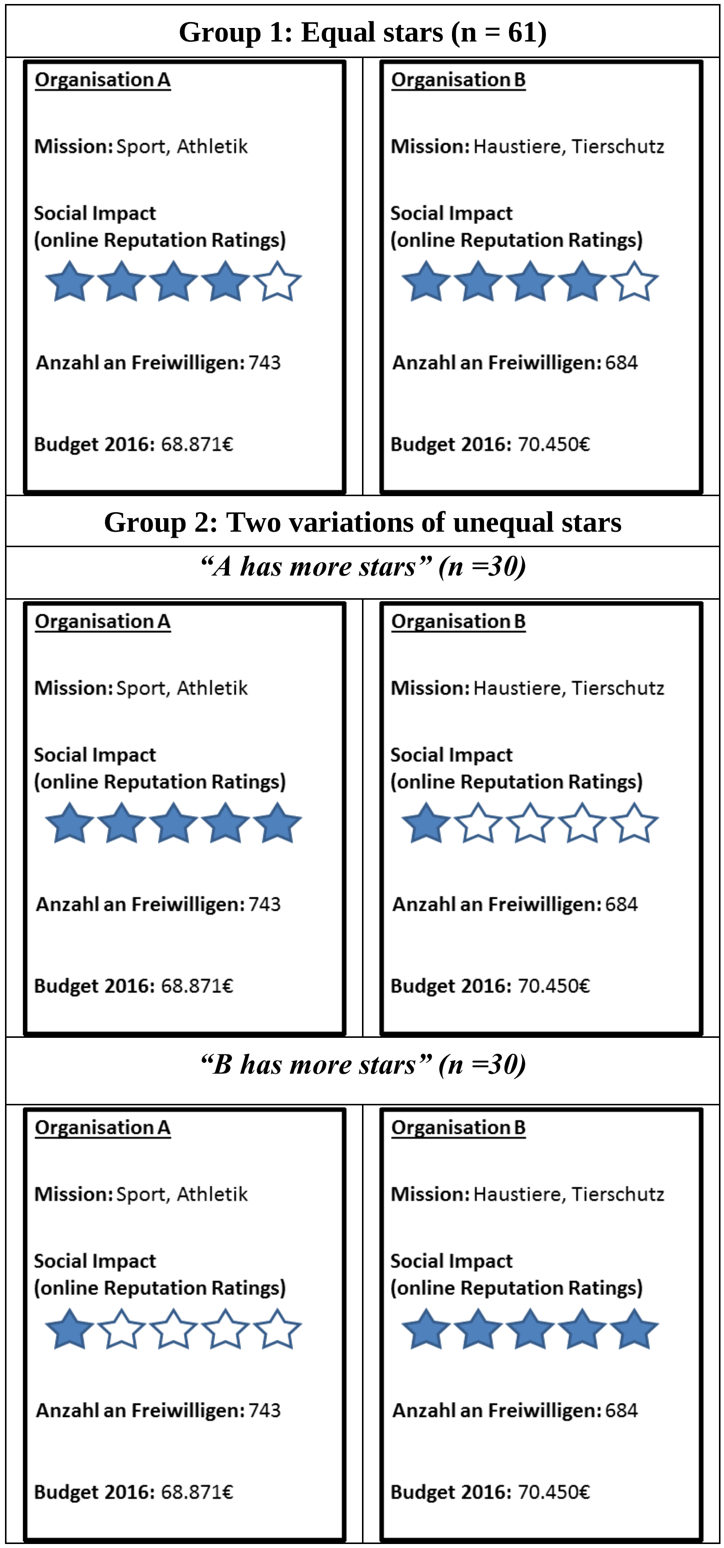
Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*(2), 260–281. https://doi.org/10.1006/obhd.2000.2909.

Zacharia, G., Moukas, A., & Maes, P. (2000). Collaborative reputation mechanisms for electronic marketplaces. *Decision Support Systems*, *29*(4), 371–388. https://doi.org/10.1016/S0167-9236(00)00084-1.

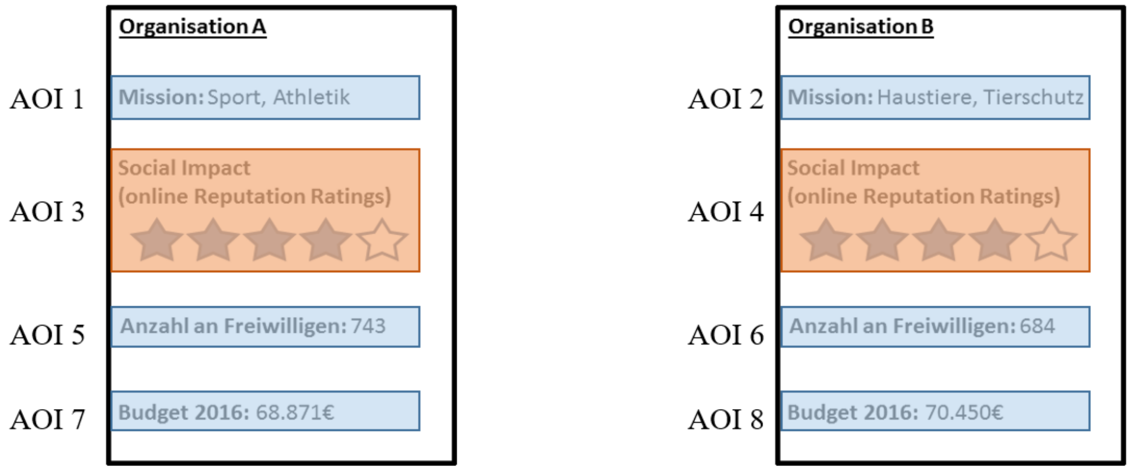
Zhang, J. (2006). The roles of players and reputation: Evidence from eBay online auctions. *Decision Support Systems*, *42*(3), 1800–1818. https://doi.org/10.1016/j.dss.2006.03.008.

Zhou, M., Dresner, M., & Windle, R. J. (2008). Online reputation systems: Design and strategic practices. *Decision Support Systems*, *44*(4), 785–797. https://doi.org/10.1016/j.dss.2007.10.001.

APPENDIX - Figure A1: Overview of organizational comparison, based on treatment groups and subgroups (equal or unequal stars in the comparison)



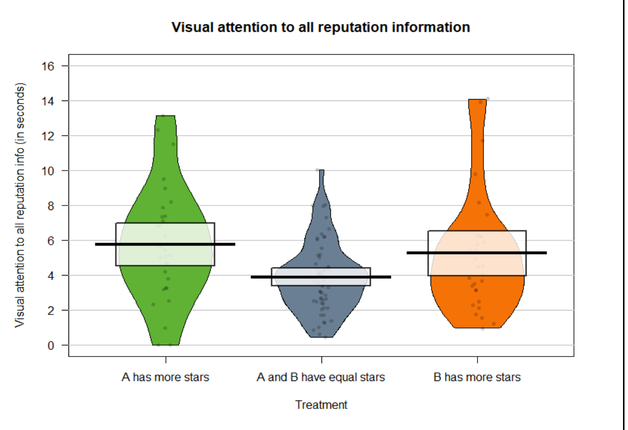
APPENDIX - Figure A2: Areas of interest (AOIs), capturing visual attention to reputation star ratings and to other available, non-reputation information.



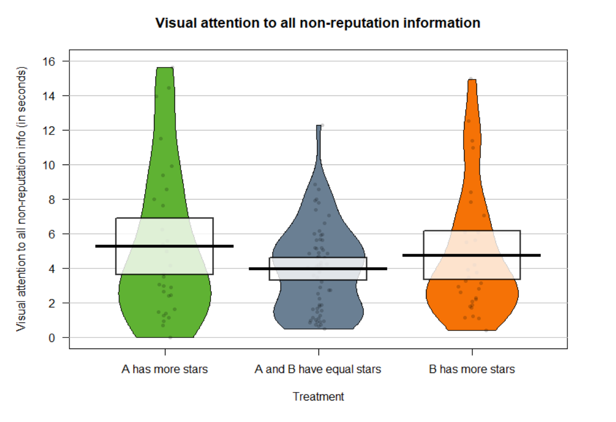
Note: The dependent variable for Hypothesis 1, i.e. visual attention to reputation star ratings, was the sum of the total fixation time on AOI 3 and AOI 4. The dependent variable for Hypothesis 2, i.e. visual attention to other available, non-reputation information, was the sum of the total fixation time on AOI 1, AOI 2, AOI 5, AOI 6, AOI 7, and AOI 8.

APPENDIX - Figure A3: Visual attention to reputation and non-reputation information: Comparing means, 95%-confidence intervals, and density per treatment group.

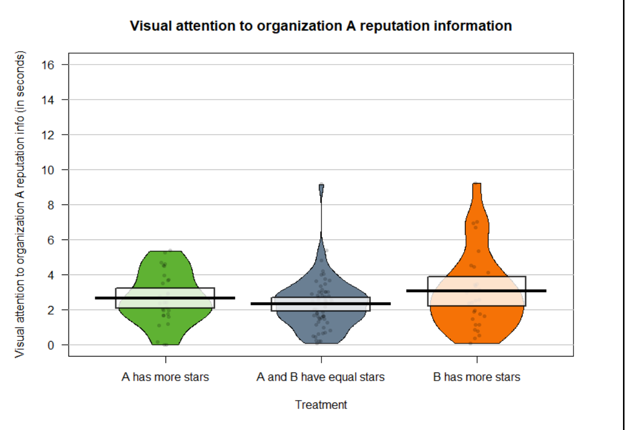
I. Visual attention to all reputation information: Comparing means, 95%-confidence intervals, and density per treatment group.



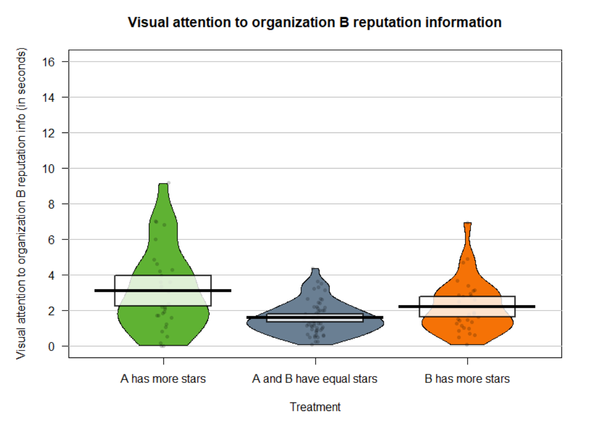
II. Visual attention to all non-reputation information: Comparing means, 95%-confidence intervals, and density per treatment group.



III. Visual attention to reputation information of Organization A: Comparing means, 95%-confidence intervals, and density per treatment group.



IV. Visual attention to reputation information of Organization B: Comparing means, 95%-confidence intervals, and density per treatment group.



Note: For Treatment group 1 (Organization A and B have equal stars), and the two subgroups of Treatment group 2 (Organization A has more stars than B; Organization B has more stars than A), group means and distributions for visual attention (in seconds; vertical axis) are reported, based on a bean plot (Kampstra, 2008; Phillips, 2017). The black horizontal line in each group reports the group mean. The box around the mean line report the 95% confidence intervals for the mean value, and the density curves/clouds give insights into the distribution of observations within each group. The wider the density cloud in a group, the more observations for the corresponding value on the vertical axis.

APPENDIX - Table A1: Descriptive statistics for visual attention, overall, and per treatment group.

		Dependent variable 1: Visual attention to reputation star ratings				Dependent variable 2: Visual attention to non-reputation information			
	N / n	mean	s.d.	min.	max.	mean	s.d.	min.	max.
All	121	4.68	2.86	0.00	14.11	4.46	3.41	0.00	15.64
Treatment Group 1 (No difference in stars)									
A and B have equal stars	61	3.88	2.01	0.43	10.03	3.94	2.56	0.50	12.30
Treatment Group 2 (A difference in stars)									
Group 2 (overall)	60	5.49	3.35	0.00	14.11	5.00	4.05	0.00	15.64
Subgroup: 'A has more stars'	*30*	*5.75*	*3.30*	*0.00*	*13.14*	*5.25*	*4.39*	*0.00*	*15.64*
Subgroup: 'B has more stars'	*30*	*5.24*	*3.44*	*0.92*	*14.11*	*4.75*	*3.74*	*0.40*	*14.97*

Note: Values are expressed in seconds, i.e. total fixation time on areas of interest (AOIs) that contained reputation-star ratings for both Organization A and B (Dependent variable 1: Hypothesis 1), or non-reputation information for both Organization A and B (Dependent variable 2: Hypothesis 2).

APPENDIX - Table A2: Complementary analysis; Explaining organizational choice based on treatment (difference in reputation stars), and visual attention to each organization.

	Prefer Organization B over Organization A			
	(1)	(2)	(3)	(4)
Constant	0.454* (0.187)	0.718** (0.273)	1.007* (0.446)	1.174* (0.469)
Treatment group 1: Org A and Org B have equal stars (*reference category*)				
Treatment group 2 (subgroup *a*): Org. A has more stars (*dummy*)		-0.851 (0.456)		-1.231* (0.554)
Treatment group 2 (subgroup *b*): Org. B has more stars (*dummy*)		-0.171 (0.467)		-0.109 (0.535)
Visual attention to the reputation-star rating of Org. A			-0.416* (0.162)	-0.503** (0.177)
Visual attention to the reputation-star rating of Org. B			0.083 (0.147)	0.234 (0.164)
Visual attention to the other available, non-reputation info of Org. A			-0.307* (0.129)	-0.272* (0.132)
Visual attention to the other available, non-reputation info of Org. B			0.582** (0.194)	0.587** (0.201)
Observations	121	121	121	121
Log Likelihood	-80.833	-79.035	-70.594	-67.853
Akaike Inf. Crit.	163.666	164.069	151.188	149.706

Method: Generalized linear logistic regression
Dependent variable is dummy-coded: “Prefer Organization B over Organization A” (= 1), “Prefer Organization A over Organization B” (= 0).
Standard errors between parentheses
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$