Association for Information Systems
AIS Electronic Library (AISeL)

Research Papers | ECIS 2019 Proceedings

5-15-2019

WHY LESS IS MORE: AN EYE TRACKING STUDY ON IDEA PRESENTATION AND ATTRIBUTE ATTENDANCE IN IDEA SELECTION

Arnold Wibmer
University of Innsbruck, arnold.wibmer@uibk.ac.at

Frederik M. Wiedmann
University of Innsbruck, frederik.wiedmann@uibk.ac.at

Isabella Seeber
University of Innsbruck, isabella.seeber@uibk.ac.at

Ronald Maier
University of Innsbruck, ronald.maier@uibk.ac.at

Recommended Citation
Wibmer, Arnold; Wiedmann, Frederik M.; Seeber, Isabella; and Maier, Ronald, (2019). "WHY LESS IS MORE: AN EYE TRACKING STUDY ON IDEA PRESENTATION AND ATTRIBUTE ATTENDANCE IN IDEA SELECTION". In Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden, June 8-14, 2019. ISBN 978-1-7336325-0-8 Research Papers.
https://aisel.aisnet.org/ecis2019_rp/41

WHY LESS IS MORE: AN EYE TRACKING STUDY ON IDEA PRESENTATION AND ATTRIBUTE ATTENDANCE IN IDEA SELECTION

Research paper

Wibmer, Arnold, University of Innsbruck, Innsbruck, Austria, arnold.wibmer@uibk.ac.at

Wiedmann, Frederik, University of Innsbruck, Innsbruck, Austria, frederik.wiedmann@uibk.ac.at

Seeber, Isabella, University of Innsbruck, Innsbruck, Austria, isabella.seeber@uibk.ac.at

Maier, Ronald, University of Innsbruck, Innsbruck, Austria, ronald.maier@uibk.ac.at

Abstract

Innovation contests often result in several hundred ideas generated. Raters have to process this huge amount of ideas that consist of attributes like idea descriptions and various types of feedback information with limited cognitive resources in order to separate good from bad ideas. It is not clear to what extent raters attend the available information during idea selection. In order to improve our understanding of how to best support raters in idea selection, this study investigated the influence of variations of the presentation mode (two versus four ideas per screen) on the attention paid to information on idea attributes using eye-tracking. We investigated attributes that refer to idea descriptions, feedback about the content of ideas (creativity score, tags) and about the community comprising the ideators and the crowd (historical success of the ideator, likes). The results of our study show that with fewer alternatives per screen, feedback attributes received more attendance, while we found no significant difference for the processing of idea descriptions. These findings provide first insights into the information-processing behaviour of raters and can inform the design of selection platforms and theory building on the effects of feedback in idea selection.

Keywords: Attendance, Attributes, Crowdsourcing, Decision making, Eye tracking, Feedback, Idea selection, Innovation contest, Presentation mode

1 Introduction

Innovation contests provide companies with the opportunity to identify exceptional ideas which, however, are usually rare and accompanied by numerous less valuable ideas (Blohm, Bretschneider, Leimeister, & Krcmar, 2010; Terwiesch & Ulrich, 2010). In addition, innovation contests tend to produce large amounts of ideas. For example, when IBM held its “Innovation Jam” in 2006, contributors generated 46,000 product ideas (Bjelland & Wood, 2008). As a result, the succeeding idea selection phase can be challenging. To deal with a huge pile of ideas, companies must deploy raters to sift through each idea and select the most promising ones. In this process, raters must carefully examine the available idea descriptions and base their decisions upon them. In addition, raters have access to *content feedback*, additional information on the idea such as tags or creativity scores, and *community feedback*, information about for example the ideator’s past performance or crowd evaluations collected during the idea generation phase of the contest. These feedback types are additional attributes of ideas besides its idea description (e.g. Svenson, 1979; Timmermans, 1993). The provision of such additional feedback information for each idea can assist raters in their decision making, but also extend the amount of presented information.

Research that was unrelated to the context of idea selection found that by presenting fewer alternatives decision makers tend to engage in a higher amount of information search (Payne, 1976; Payne & Braunstein, 1978) and, therefore, visually attend to more attributes (Lohse & Johnson, 1996). To the best of our knowledge, no study has yet explored the link between idea presentation and its association to the attendance of idea attributes in idea selection. As introduced above, submissions to innovation contests can include various attributes in form of community feedback or content feedback. The degree to which raters consider such attributes in the decision-making process can affect the selection strategy. It is not clear whether different presentation modes of ideas alter the attendance to idea attributes such as idea descriptions and feedback Hence, we do not have sufficient understanding of how to best design IT tools to support raters in their idea selection decisions. We therefore investigate the research question: *How do variations in idea presentation influence raters' attendance to idea attributes?*

The goal of this study is to shed light onto raters' attendance to idea attributes when presented with fewer versus more ideas. We understand attendance to attributes as visual attendance (Balcombe, Fraser, & McSorley, 2015) describing the degree to which an attribute was looked at and visually examined by the decision maker during the idea selection task. We build on the concept of the adaptive decision maker to explain that information search behavior regarding feedback attributes is contingent on the task environment (Bettman, Luce, & Payne, 1998; Payne, Bettman, Coupey, & Johnson, 1992; Payne, Bettman, & Johnson, 1988, 1993), in our case the presentation mode. To test our hypotheses, we designed a laboratory experiment, manipulated the presentation mode (2 ideas at a time vs. 4 ideas at a time), and collected data on attribute attendance with eye-tracking.

We structured our paper in the following way: The background section presents the idea of an adaptive decision maker and reviews consequences for the amount of information that is searched (selective information search). Subsequently, we develop our hypotheses and present our research model. In the method section, we describe our conducted eye tracking experiment and operationalize our measures. The results section reports our analysis and findings from presenting either two or four ideas at a time. Finally, we discuss our contributions for both researchers and practitioners, look at our findings through the perspective of digital nudging and list limitations of our experiment.

2 Background

2.1 Adaptive Decision Maker

The term adaptive decision maker was motivated by the idea of individuals constructing their decision strategies and adapting them to different situations (Payne et al., 1993). As stated by Simon (1955), decision makers have limited cognitive resources and are aware of their cognitive capacities. Due to their cognitive restrictions, decision makers deliberately include simplification strategies into their decision model and adopt contingent decision behaviors. Decision behaviors are contingent as decision strategies are developed on the spot out of fragments from memory instead of taken from a master list of decision strategies. They are therefore highly sensitive to the local problem structure (Bettman et al., 1998; Payne et al., 1992). Some decision strategies require more effort than others as more available information is considered, which can help ensuring a high choice accuracy. Other decision strategies require less effort by considering less information and still achieve a comparably high choice accuracy in certain choice environments (Johnson & Payne, 1985). A decision maker might a priori intend to apply an effortful decision strategy to a decision problem considering all available information. However, the characteristics of the choice environment can lead to a switch to a less effortful strategy (Payne et al., 1992). Adaptations of the applied decision strategy to a choice environment have been found in several studies (e.g. Biggs, Bedard, Gaber, & Linsmeier, 1985; Kerstholt, 1992; Klein & Yadav, 1989; Payne et al., 1988; Russo & Dosher, 1983). The used decision strategy seems to depend on task factors and context factors. Task factors reflect general characteristics of a decision problem, like the number of available alternatives to a decision maker, different response modes or repetitive decision problems with several rounds. Context factors relate to different values of alternatives in the decision problem. In an idea contest, such values of context factors are attributes such as likes or ratings. The values of one alternative could be similar or different to the attribute values of another alternative. They can even dominate or be dominated by the values of other alternatives. The values are linked to the effort a decision maker faces when making a choice. For example, it should be easier to select an idea with many likes and good ratings when all other ideas have no likes and poor ratings. Hence, Both task and context factors influence the construction and application of distinct decision strategies (Klein & Yadav, 1989; Payne et al., 1992).

2.2 Selective Information Search

Decision strategies do not necessarily include the processing of all available information. Some decision problems are more complex due to certain task or context factors and decision makers rely on a smaller amount of information. Hence, some decision strategies allow selective processing of information. Selectivity can appear in two ways. If a decision strategy is selective across attributes, the decision maker considers some attributes of an alternative, but stays consistent across alternatives. In contrast, if a decision strategy is selective across alternatives, the decision maker skips attributes only for some alternatives (Bettman et al., 1998; Riedl, Brandstätter, & Roithmayr, 2008).

The number of available alternatives is a task factor that has received wide attention in research. Pioneering studies on information acquisition behavior found more alternatives to be related to a lower proportion of searched information, i.e. with more alternatives a lower percentage of attributes was examined (Payne, 1976; Payne & Braunstein, 1978). In this context, evidence was found for decision makers switching from a one-stage to a two-stage decision strategy. Instead of deciding in one go, decision makers were found to apply an initial decision strategy of screening followed by a second decision strategy to actually make a choice. A frequent two-stage approach was for example to begin by making a pre-selection based on cut-off values and to make the final choice in favor of the alternative with the most dominating attributes (Olshavsky, 1979). Apparently, more available information from both more alternatives and more attributes results in a smaller proportion of information acquisition (Biggs et al., 1985; Svenson, 1979). Consumer research on assortment size added new insights to this relationship. If more product alternatives are available, consumers still examine almost all alternatives, but become selective regarding less important (secondary) attributes. While search for primary attributes remains consistent, consumers examine fewer secondary attributes (Dörnyei, Krystallis, & Chrysochou, 2017). This effect has been explained with higher information load for decision makers when offering more available alternatives and attributes (Jacoby, 1984; Malhotra, 1982). The relationship between more information load through more available alternatives had further been found for capital investment decisions (Swain & Haka, 2000) and health care insurance decisions (Schram & Sonnemans, 2011). In fact, Hensher (2006) found decision makers to ignore more attributes in their decisions under increasing information load.

3 Hypotheses

The principle assumption that attendance to attributes decreases when the number of alternatives and attributes increases due to information load (Hensher, 2006; Schram & Sonnemans, 2011; Swain & Haka, 2000) should also be applicable to the context of idea selection. On the one hand, decision-makers can ignore information during selection in order to cope with information overload and therefore improve decision quality. On the other hand, some attributes might contain relevant information on whether or not an idea is promising, for example community feedback on why and how the implementation of an idea might satisfy specific customer needs and thus, when ignored, might reduce decision quality.

As the number of alternatives (ideas) presented at a time increases, the required cognitive effort to make a selection rises (Bettman, Johnson, & Payne, 1990; Payne, Bettman, & Johnson, 1988). Due to the cognitive restrictions of human decision makers (Miller, 1956), decision makers should adapt the decision strategy to a less effortful one resulting in a lower amount of information search (Kerstholt, 1992; Lohse & Johnson, 1996; Payne & Braunstein, 1978; Swain & Haka, 2000) for idea descriptions, idea content feedback and idea community feedback.

In idea selection, a key element is the idea description. Human decision makers can read through the description and interpret its content beyond its words or grammatical structure (Hoornaert, Ballings, Malthouse, & Van den Poel, 2017). Decision makers are expected to base their assessment of the quality on the description to infer if an idea is relevant, feasible, novel and elaborated, which represent common evaluation criteria (Dean, Hender, Rodgers, & Santanen, 2006). In addition, idea descriptions tend to encompass more words than other ideas attributes, such as community feedback or content feedback. Considering this, one would assume that completely ignoring an idea description should not be very likely. However, the length of an idea description makes reading and processing this attribute a time-consuming process and might demand high cognitive resources when ideas are unfocused and not optimally elaborated (Beretta, 2018).

We expect the requirements for cognitive processing to rise on a page with more ideas presented at a time since there is more information to process. In contrast, the processing capacity of a decision maker still remains the same. Due to the required cognitive processing of presented information that exceed one's own cognitive capacity, we argue that decision makers adapt to a less effortful and more selective strategy. Taking this perspective, research found that decision makers request less information to make a choice when more alternatives are presented (Kerstholt, 1992). It could be that decision makers distribute their limited processing capacity between the presented idea descriptions, resulting in less processing per idea description when more ideas are presented. Hence, presenting fewer alternatives at a time should make it more likely that decision makers read and process idea description in-depth as they distribute their limited cognitive resources among fewer idea descriptions."

> *H1: Presenting fewer alternatives at a time leads to more intensive processing of idea descriptions than presenting more alternatives.*

A variety of information might be available beyond the idea descriptions in order to help raters to select the most promising ideas. Hoornaert et al. (Hoornaert et al., 2017) provide a framework that classifies the sources of information available for idea selection decisions into content-, contributor- and crowd-based information. Concerning, content-based information, recent developments in text mining allow to generate *content feedback*. Such text mining approaches can either provide additional feedback on idea description characteristics or produce key words as categorization. For example, Toubia & Netzer (2017) developed a text-mining algorithm, which assessed familiarity and novelty of an idea based on its idea description. According to their algorithm, creative ideas should exist where familiarity and novelty is balanced.

Other sources of feedback on ideas can origin from innovation communities comprising contributor-based and crowd-based information, which allows producing *community feedback*. A contributor's history of successful ideas, for example, has been found to be an indicator of idea quality as some people generate better ideas than others (Girotra, Terwiesch, & Ulrich, 2010). Feedback from the crowd, often labeled as the wisdom of the crowd (Surowiecki, 2004), contains the opinion of community members about an idea, e.g., using likes or ratings. A crowd is likely to consist of potential users and, if composed well, can serve as a proxy for expert ratings capable of identifying good ideas (Kornish & Ulrich, 2014; Magnusson, Wästlund, & Netz, 2016).

Both kinds of feedback, i.e. content feedback and community feedback, provide additional information that a decision maker could consider when selecting ideas. Given that more attributes and alternatives lead to a higher information load (Hensher, 2006; Swain & Haka, 2000), participants' search strategy becomes less exhaustive and less systematic and attributes are ignored. Similarly, research on online purchase behavior found customers to pay less attention to information if the information load was high (Sicilia & Ruiz, 2010). Similarly to our argumentation for the idea description, presenting more ideas at a time is likely to result in more cognitive effort (Bettman et al., 1990; Payne et al., 1988). This should trigger decision makers to adapt to a decision strategy that is less exhausting by skipping or ignoring attributes. Thus, we assume that presenting more ideas at a time will result in a similar effect on content feedback and community feedback with receiving less attention by the decision maker.

H2: Presenting fewer alternatives at a time leads to higher attendance of content feedback than presenting more ideas.

H3: Presenting fewer alternatives at a time leads to higher attendance of community feedback than presenting four alternatives.

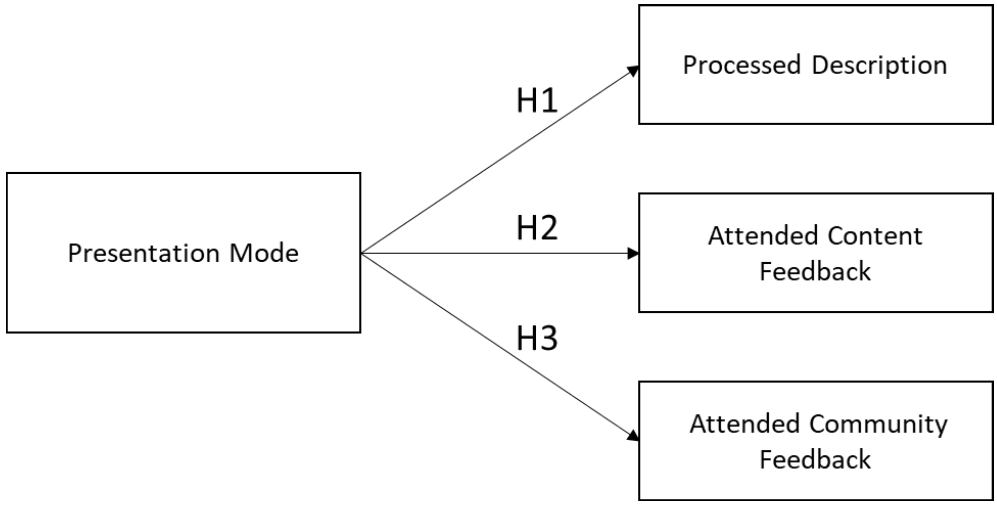

Figure 1. Research Model

4 Method

In order to test our hypotheses, we designed a laboratory experiment using eye-tracking methods in which participants had to evaluate 32 ideas.

4.1 Treatment and Eye Tracking

We manipulated the presentation mode with the number of alternative ideas presented on a screen at a given time. In one treatment condition participants saw 16 screens of two ideas each. In the other treatment condition, they saw 8 screens of four ideas each. We defined non-overlapping areas of interest (AOIs) to collect eye-tracking data on the attributes of interest: *idea description* (incl. title of idea), *community feedback* (historical idea score and number of likes), as well as *content feedback* (creativity score and tags) on each screen. The AOIs allowed us to track gazes and count the fixations on the area spanned by the AOI. We positioned the AOIs as recommended by the supplier Tobii with an error frame of 1 degree corresponding to 32 pixels in all directions. This should counteract errors in calibration. We had to cut 12 pixels from the lower part of the AOIs to avoid overlapping with other AOIs due to the limited screen space (the display size was 24 inch with a resolution of 1920x1080). We collected the data using a Tobii Pro X3-120 eye tracker with sample rate of 120 Hz.

We used different thresholds for fixation times for idea description and feedback attributes. Due to the fact, that the idea description resembles a reading task and should be related with more in-depth elaboration of content, we used a fixation time of 251ms. Past research suggests that the average reading speed of adults is between 60 ms and 500 ms, being 250 ms on average (Liversedge & Findlay, 2000). At the same time, Glöckner and Herbold (2011) adopted fixations of higher than 250 to investigate medium and long information processing. Fixations on the community and content feedback attributes were identified using Tobii's standard value of 60 ms as minimum fixation time (Olsen, 2012) because also Glöckner & Herbold (2011) suggest fixation times lower than 250 ms as indicator for scanning attributes. Figure 2 shows a screenshot of the AOIs of the treatment condition two ideas per screen for screen number S01 with the AOIs for idea description (AOI_descr), creativity score (AOI_cs), likes (AOI_like), historical idea score (AOI_his) and tags (AOI_tag).

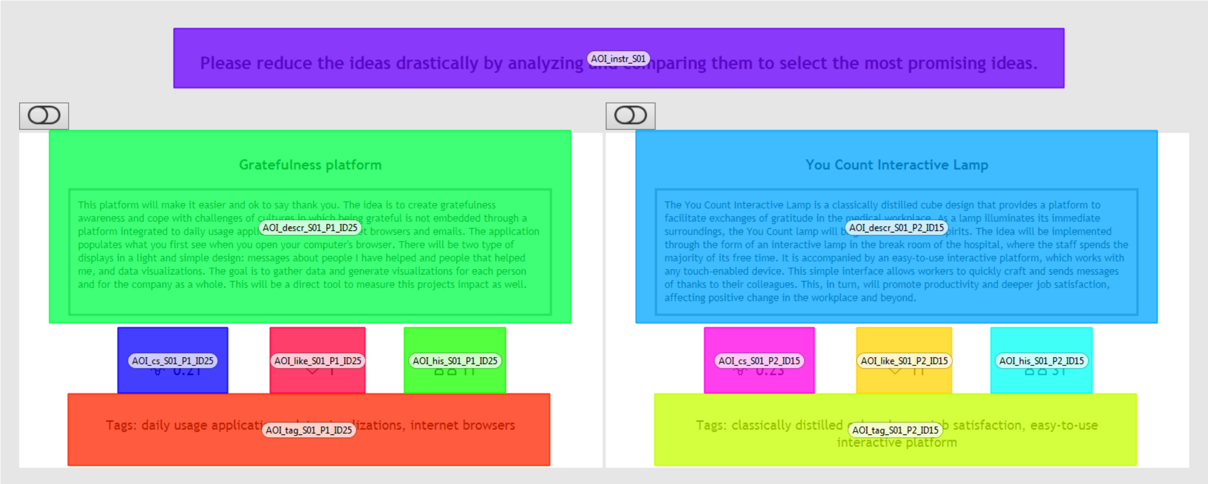

Figure 2. Selection task with two alternatives per screen and AOIs

4.2 Idea Set

We initially drew a stratified sample of 40 ideas from a real online idea competition ("OpenIDEO") that focused on gratitude at the workplace. Since the original descriptions of the ideas were too long, they had to be reworked to control for idea length. In the rewriting process, meaningful and summarizing sentences were copied out of the original idea description and put together so that the "What" and the "How" of each idea was sufficiently described. The maximum length of the reworked idea descriptions was 130 words.

To ensure that our stratified sample included good ideas, we asked four experts with background in Human Resources to evaluate this set of 40 ideas based on the four criteria (feasibility, novelty, elaborateness and relevance). Building on Blohm et al., (2010) who suggested that 10-30% ideas can be considered good ideas in innovation contest, we considered the best 25% or 10 ideas as good ideas. We dismissed eight ideas so that the final idea sample consisted of 32 ideas. We calculated Krippendorff's Alpha (Hayes & Krippendorff, 2007) on the four criteria to estimate the interrater agreement between the experts. The results can be viewed in Table 1.

Criteria	**Krippendorff's Alpha on original rating**	**Krippendorff's Alpha on z-transformed values**
Elaborateness	0.090*	0.220**
Feasibility	0.177*	0.249**
Novelty	0.339**	0.431***
Relevance	0.142*	0.192*

Table 1. Interrater agreement with Krippendorff's Alpha z-transformed rating scales of the criteria. Agreement is evaluated with the levels of (Landis & Koch, 1977, p. 165): * Slight Agreement (.00 -.20), **Fair Agreement (.21-.40), ***Moderate Agreement (.41-.60), ****Substantial Agreement (.61-.80).

To determine the agreement among raters, we assessed Krippendoff's Alpha on original and z-transformed measures (see Table 1), which is similar to Magnusson et al. (Magnusson, Netz, & Wästlund, 2014). After the z-transformation, we observed a slight agreement for relevance, a fair agreement for elaborateness as well as feasibility and a moderate agreement regarding novelty. We therefore calculated a mean score of z-transformed values over all criteria.

Blohm et al., (2010) identified that in a set of generated ideas, around 10-30% can be considered good ideas. In this study, we chose the upper boundary and ensured that the final idea set contained 30% good ideas, which equaled 10 ideas. It should be noted that this gold standard assessment was important to make sure that the experiment mirrored a real-world scenario, which would also include a small number of good ideas and a larger amount of less valuable ideas.

4.3 Measures

Processed description refers to what extent participants read the *idea description*. The variable was operationalized with the number of fixations on the AOIs defined on the idea description. This means a fixation was counted as soon as a participant fixated on the idea description longer than the threshold of 251 ms (Glöckner & Herbold, 2011). To receive an overall number of processed description per participant, we summed up the counts across all AOIs defined for idea descriptions.

Attendance to content feedback describes to what extent participants visually attended the AOIs we defined for the two attributes creativity score and tags. The *creativity score* assesses how creative an idea is by balancing familiarity and novelty. We adopted the text-mining algorithm by Toubia & Netzer (2017) who argue that a creative idea would be one that balances familiarity and novelty and therefore has a creativity score, which is close to 0. *Tags* provided feedback to which conceptual category the idea could fit. We used the IBM artificial intelligence service "Natural Language Understanding" to extract key words from the idea description. Words that equaled "idea" or "gratitude" were excluded from the list. For each idea, we selected those three key words from the remaining list with the highest relevance score provided by the service. We operationalized the attendance to content feedback by calculating the ratio of the number of content feedback attributes (AOIs for creativity score and tags) fixated by a participant at least once divided by all content feedback attributes presented to the participants. For example, a value of 0.75 for attendance to content feedback means a participant fixated 75% of the attributes of content feedback and skipped the remaining 25%.

Attendance to community feedback describes to what extent participants visually attended the AOIs we defined for the two attributes likes and historical idea score. *Likes* measures the number of "loves" that registered users provided to ideas. We took the original counts of likes from the OpenIdeo platform, which ranged from 1 to 20 with a mean of 6.80 for good ideas and from 1 to 20 with a mean of 5.41 for bad ideas. The *historical idea score* measures the past appreciation of the ideator's contributions to the platform, such as sharing an idea, adding a post or an evaluation of others' ideas, which was labeled as design quotient on the OpenIdeo platform. The span of the score was from 11 to 100 with a mean of 39.50 for good ideas and from 11 to 101 with a mean of 33.91 for bad ideas. We operationalized the attendance to community feedback by calculating the ratio of the number of community feedback attributes (AOIs for likes and historical idea score) fixated by a participant at least once divided by all community feedback attributes presented to the participants.

Control variables: We collected information on *gender*, *contest experience*, and *English proficiency* to control for individual differences that may influence an individual's performance in idea selection. Moreover, we control for *regulatory focus,* which we manipulated by priming participants for their strategy in decision making according to Chernev (2004) to have either a promotion focus, i.e. search for the best ideas in the set, or a prevention focus, i.e. prevent bad ideas from being declared as good.

4.4 Participants

Thirty-one graduate students (20 males and 11 females) from a European University participated in the experiment. Participants' age ranged between 22 and 29 (mean age = 25.23, SD = 2.109). We advertised the study in a Master course on "Research Methods". Twenty students participated and received class credits. We also invited graduate students writing their Master theses in Information Systems and those 11 students participated without further incentives. Of the 31 participants, 18 already had work experience, six had participated in an innovation contest before and one had prior experience as jury member of an innovation contest. Six participants wore glasses, which had no negative effect on the eye tracking data collection procedure as we provided cloths for cleaning. Participation was voluntary, students were informed about potential risks, and informed consent was collected before the experiment started. The participants could quit the experiment at any time.

4.5 Experiment Procedure

Prior to the experiment, the participants were randomly assigned to one of the experimental groups by a computerized random number generator. Each experiment was conducted with one participant at a time. Participants sat in front of a computer screen with a mouse and key board placed in front of them. The experimenter was positioned at a table next to the participant in front of a laptop. After a brief overview over the procedure of the experiment, participants were instructed for two warm-up activities prior to the eye tracking. Both activities resembled a priming procedure of Chernev (2004) by relying on a combination of two traditionally used priming techniques. In the first activity, participants were instructed to write about their hopes and aspirations or duties and obligations as graduate students. Writing about hopes and aspirations is associated with priming towards a promotion focus, while writing about duties and obligations is associated with priming towards a prevention focus (Freitas, Higgins, Freitas, & Higgins, 2002; Higgins, Roney, Crowe, & Hymes, 1994). The second activity included an adapted version of the paper and pencil maze used by Friedman & Forster (2001). A cartoon mouse was placed inside a maze with a mouse hole waiting at the exit. In the promotion focus condition, a promotion cue in form of a piece of cheese was placed at the entrance of the mouse hole and participants were instructed to find the way for the mouse to get the cheese. In contrast, the prevention focus condition contained a cartoon eagle approaching the mouse. Participants were instructed to find a way to escape the eagle. Subsequently, the experimenter supported participants to prepare for the eye tracking and to keep a distance of 60-65 cm between the participants' eyes and the eye tracker. After a successful calibration in which participants had to follow a red dot across the screen, a website-based idea selection task opened at the screen. From that point on, participants eye gazes were recorded. The idea selection task was prefaced by an exemplary idea form without any content. Participants were instructed with the words "*Please reduce the ideas drastically by analyzing and comparing them to select the most promising ideas*". In the idea selection task, participants clicked through eight respectively sixteen screens with either four or two ideas each. During the task, they had no opportunity to go back to previous screens and change their selection decisions. Subsequently, the eye tracking was stopped and participants filled out a survey collecting control variables for the analysis. The experiment ended with a debriefing and check for suspicion. The experiment was conducted between May and July 2018.

4.6 Assumption Tests and Analysis Procedure

We used IBM SPSS Statistics 25 to analyze the collected data. Prior to running a multivariate analysis of covariance (MANCOVA) to test our hypotheses, we tested if our data violates any of the statistical assumptions. We performed univariate outlier analysis using z-scores and identified one case, which showed significant deviations in four measured items. A closer examination of the eye-tracking data and log revealed that the participant's calibration result was borderline, which questioned the reliability of the collected data for this case. Hence, we excluded this case and repeated the analysis. There were no further outliers that exceeded the recommended threshold of +/-2.5 (Hair, Black, Babin, & Anderson, 2014). We tested the assumption of multivariate normality with the Shapiro-Wilk test and inspecting histograms. The Shapiro-Wilk test was insignificant for the variables "description" and "content feedback", but significant for the variable "community feedback" ($p < 0.05$) suggesting a violation of normality. We tested the assumption of homoskedasticity with the Box M and Levene's test. All variables met the test assumptions ($p > 0.05$ (Hair et al., 2014), which suggests that homogeneity of variance can be assumed. Therefore, we proceeded with hypotheses testing using MANCOVA. We also performed a Mann-Whitney-U test to investigate hypothesis H3 as a robustness check for the construct "community feedback" which did not fulfill the normality assumption.

5 Results

We performed a MANCOVA to assess if the treatment variable presentation mode has a multivariate effect on idea descriptions, content feedback, community feedback as well as description given the control variables regulatory focus, contest experience, English proficiency and gender. There exists a significant multivariate effect for presentation mode (Pillai's trace = 0.300, $F(3,22) = 3.144$, $p = .046$, $\eta^2 =$

0.300, power = 0.648) that has a large effect size (Cohen, 1988). We present the descriptive statistics in Table 2 and the results of the MANCOVA in Table 3. The descriptive statistics in Table 2 shows the average number of fixations on the idea description and the average attendance (in %) to the content feedback as well as to the community feedback. Table 3 describes the statistical significance of the individual measures Regulatory Focus, Gender, Contest Experience, English Proficiency and Presentation Mode. Given the significant multivariate difference for presentation mode, we followed up univariate ANCOVAs to test hypotheses 1 through 3.

Variable	**Two alternatives per screen (N=16)**	**Four alternatives per screen (N=14)**
	Mean (SD)	**Mean (SD)**
Processed Description (Fixation Counts)	1097.44 (416.42)	850.29 (366.91)
Content feedback (Attendance in %)	0.6807 (0.2055)	0.5078 (0.1937)
Creativity score	0.7441 (0.2187)	0.5603 (0.2633)
Tags	0.6172 (0.2467)	0.4554 (0.1842)
Community feedback (Attendance in %)	0.7480 (0.1630)	0.5547 (0.2481)
Likes	0.7148 (0.1847)	0.5647 (0.3413)
Historical idea score	0.7813 (0.1879)	0.5446 (0.2292)

Table 2. Descriptive statistics

Hypothesis H1 suggests that fewer alternatives per screen (2 ideas) lead to more fixations on the *idea description*. Our data shows that participants that were presented with two ideas showed a higher tendency to attend the description ($M = 1097.44$, $SD = 416.42$) than participants that were presented with four ideas ($M = 850.29$, $SD = 366.91$). However, this difference was not significant ($F(1, 24)= 2.361$, $p > 0.05$) and therefore, H1 is not supported.

Hypothesis H2 suggests that fewer alternatives per screen (2 ideas) lead to a higher attendance to *content feedback*. Our data shows that participants that were presented with two ideas attended to more content feedback ($M = 0.6807$, $SD = 0.2055$) than participants that were presented with four ideas. This difference was significant, which supports H2, and shows a large effect size ($F(1, 24)= 6.074$, $p < 0.05$, $\eta^2 = 0.202$).

Hypothesis H3 suggests that fewer alternatives per screen (2 ideas) lead to a higher attention to *community feedback*. Due to the fact, that the Shapiro Wilk test indicated non-normal distribution of data, we followed up with a Mann-Whitney-U -test as a robustness check. The test indicates that participants show a significantly higher attendance to *community feedback*, if 2 ideas were presented per screen ($U = 57.000$, $p < 0.05$). Therefore, H3 is supported.

	Pillai's trace	**F**	**Sig.**	**Effect size (partial eta squared)**
Intercept	0.242	2.339	0.101	0.242
Regulatory Focus	0.091	0.733	0.543	0.091
Gender	0.104	0.854	0.480	0.104
Contest Experience	0.286	2.930	0.056	0.286
English proficiency	0.054	0.421	0.739	0.054
Presentation Mode	0.300	3.144	0.046	0.300

Table 3. Levels of the multivariate test

6 Discussion and Implications

The aim of this work was to investigate how information is attended in idea selection tasks. To do that we examined different *modes of idea presentation (two versus four ideas)* and its effects on the *attendance of idea descriptions, content feedback* and *community feedback* using eye-tracking. Our findings contribute to the literature on idea evaluation and assortment sizes in two ways.

First, we found that participants attend to more feedback attributes when they are presented with fewer ideas. The feedback attributes we studied cover content-, contributor- and crowd-based information that were found to typically characterize information in idea selection environments (Hoornaert et al., 2017). Our findings suggest that feedback attributes attract more attendance if participants saw fewer ideas per screen. Thus, our results confirm theory that argued for a lower proportion of information sought with more alternatives (Payne, 1976; Payne & Braunstein, 1978). Recent studies on consumer purchasing decisions also confirmed the systematic impact of assortment size and attribute quantity on decision-making and choice outcome (Dörnyei et al., 2017). Our findings extend theory by changing the size of the subset via presentation mode and keeping the assortment size the same. The change of the presentation mode, in turn, influences the behavior of the decision makers.

Second, we found that in the cases of two ideas per screen, the participants on average looked more frequently at the descriptions (1097.44 vs. 850.29 fixation counts), yet this difference was not significant. This result may be related to the difference between the two modes with respect to the variation of the number of alternatives presented per screen. The effect of an increasingly demanding processing of the provided information with an increasing number of alternatives might be larger with more than four alternatives per screen and the corresponding larger difference in information load between the manipulations. However, the tracking of eye movements is not (yet) precise enough to allow the presentation of more than four descriptions of alternatives plus their attributes to the participants in a way so that AOIs can be set sufficiently apart from each other in order for the eye tracker to classify correctly which AOIs the participants look at. Moreover, a rough examination of our results regarding fixation duration indicates that the processing of a single idea description was deeper when fewer ideas were presented. Participants fixated on an idea on average 12.44 seconds in the 2 ideas per screen treatment compared to 9.78 seconds in the 4 ideas per screen treatment. Based on an exploratory analysis of some reading patterns, we gained the impression that some participants adopted an alternative-based processing and read attentively the first few sentences of the description, which later ended in a jumpy information admission by skipping some parts of the paragraphs. Intended pattern analysis of the idea description may provide interesting insights into the optimal length of description texts as well as the extent to which the participants compared available attributes together.

Our contributions have also implications for research and practice. This study is, to the best of the authors' knowledge, the first that provides objective measurement data on attribute attendance in idea selection. Eye tracking is a very precise method of data collection (Glaholt & Reingold, 2011) and provides a reliable and trustworthy method to operationalize attendance on attributes without the possibility to be manipulated by the participants. Research can benefit from our findings, because our findings show that the relevance of feedback information in choice tasks might be misleading. For example, related research showed that the number of likes or idea ratings might be important indicators for high quality ideas (Armisen, Majchrzak, & Brunswicker, 2016; Gross, 2017). Görzen and Kundisch (Görzen & Kundisch, 2017) manipulated crowd feedback to investigate if the presentation of ratings could have an anchoring effect by assessing the variance of evaluations. In their experiment, a single crowd worker evaluated 10 ideas out of 80; a closer explanation of the presentation mode is not given. Their findings show that the visualization of ratings explained a high share of the variance of rater evaluations in such that the variance is lower when evaluating good ideas and higher when evaluating bad ideas. In extension of these findings, our study suggests that raters did attend to community feedback only about half of the time when participants saw four ideas on their screen; attendance rose to about 74% when showing two ideas at a time. Hence, researchers who study the effects of feedback by designing treatments with varying degrees of feedback should be aware that their manipulation might not be considered by decision-makers as one might hope and our findings suggest that such attendance is dependent on the idea presentation mode.

Our findings also contribute to literature on digital nudging as the two idea presentation modes represent user-interface design elements that alter people's behavior in an online environment where they are supposed to choose among alternatives (Weinmann, Schneider, & Brocke, 2016). We theorized that presenting fewer alternatives at a time will result in more attention to feedback attributes and we could demonstrate that attendance was in fact higher for the two ideas per screen condition. Depending on the design of digital nudges, one can foster or hamper psychological effects, such as attentional collapse or anchoring (Mirsch, Lehrer, & Jung, 2017) people on feedback. It could be that our digital nudge of idea presentation and the presentation of feedback anchored decision makers on the ideas. Related research found that the feedback from others might be considered more when the to-be-evaluated idea is a good idea. In contrast, low quality ideas represent a less complex evaluation situation and therefore, the feedback anchor has less influence on the decision making process (Gorzen and Kundisch, 2017). Ten of the thirty ideas in our idea set were considered good ideas given the gold standard assessment. These good ideas could have driven the found differences in feedback attendance as decision makers used feedback as their anchor in their decision making. Hence, researchers that want to replicate our study should also consider the potential unintended psychological effect of the idea presentation mode.

Our findings also offer design recommendations to practitioners for the development of idea selection platforms. Results demonstrated that raters might spend only a short time on reading ideas and the presentation mode could influence to what extent they attend to feedback attributes. When contest managers would like to nudge their raters into more accurate information processing, the presentation mode pairwise comparison is recommended. If contest managers are interested in quick reduction, presenting many alternatives most likely forces participants into a less effortful search strategy in which certain information gets ignored.

7 Limitations and Future Research

As with any study involving experimental methods, our study has some limitations. The underlying study has a modest sample size of 31 participants. Nevertheless, it should be considered that eye-tracking provides reliable data, but is complex to conduct and time-consuming to analyze. Moreover, the analyses showed large effect sizes but limited power (< 0.8) (Hair et al., 2014). In our future research, we intend to increase the sample size to further validate our findings.

Moreover, the present study argued theoretically that decision makers in the 2 idea per screen treatment adapt to more effortful decision strategies and hence more attendance to idea attributes than decision makers in the 4 idea per screen treatment. Eye tracking would allow to also follow gaze paths on each screen and hence provide empirical insights. To further investigate the switching behavior of adaptive decision makers, it would be interesting to assess such information search patterns (Russo & Dosher, 1983) across screens to understand to what extent decision makers engage in alternative-based and attribute-based decision strategies. Collecting eye-tracking data on pupil dilation as an indicator of information load, would further allow to confirm the basic assumption that the reason for ignoring attributes and alternatives is due to high information load.We had to adjust the lower bound of the AOIs for 12 pixels to avoid overlaps. In any case, if the eye tracker is precisely calibrated, there will be no fixations in the region near the edge of the AOI. Therefore, we do not consider the marginal cutting of the lower edge of the AOI to be critical. This is also the reason why we could not implement a higher number than 4 ideas per screen, because this would have resulted in insufficient reliability of the eye-tracking data due to overlapping areas of interest (AOI).

Moreover, we engaged students as surrogates for raters to evaluate gratitude at work. We deliberately chose a contest which does not require profound (technical) domain knowledge, neither to generate nor to evaluate the ideas. Therefore, this was not considered as critical. Our study may not be generalizable to idea selection studies where people need to have profound domain knowledge. Therefore, future research might validate our findings in contest domains that require varying levels of domain knowledge required or need to rely on professional raters only. Lastly, it must be noted that, against our expectations, the creativity score we generated out of the idea description on average was only marginally closer to zero for good ideas than for bad ideas. Creativity scores for the selected ideas ranged from 0.13 to 0.36 with a mean value of 0.26 for good ideas and 0.17 to 0.34 with a mean of 0.24 for bad ideas. Thus, the creativity score may have not been a reliable indicator to identify promising ideas. We cannot eliminate the possibility that some participants started ignoring this attribute. Future research should investigate the relationship between community feedback attendance and selected idea quality, given certain idea presentation modes, to help falsify the feedback-idea quality hypothesis.

References

Armisen, A., Majchrzak, A., & Brunswicker, S. (2016). Formative and Summative Feedback in Solution Generation: The Role of Community and Decision Support System in Open Source Software. *ICIS 2016 Proceedings*.

Balcombe, K., Fraser, I., & McSorley, E. (2015). Visual Attention and Attribute Attendance in Multi-Attribute Choice Experiments. *Journal of Applied Econometrics*, *30*(3), 447–467. https://doi.org/10.1002/jae.2383

Beretta, M. (2018). Idea Selection in Web-Enabled Ideation Systems. *Journal of Product Innovation Management*, *00*(00). https://doi.org/10.1111/jpim.12439

Bettman, J. R., Johnson, E. J., & Payne, J. W. (1990). A Componential Analysis of Cognitive Effort in Choice. *Organizational Behavior and Human Decision Processes*, *45*(1), 111–139.

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research*, *25*(3), 187–217. https://doi.org/10.1086/209535

Biggs, S. F., Bedard, J. C., Gaber, B. G., & Linsmeier, T. J. (1985). The effects of task size and similarity on the decision behavior of bank loan officers. *Management Science*, *31*(8), 970–987. https://doi.org/10.1287/mnsc.31.8.970

Bjelland, O. M., & Wood, R. C. (2008). An Inside View of IBM's 'Innovation Jam.' *MIT Sloan Management Review*, *50*, 32–40.

Blohm, I., Bretschneider, U., Leimeister, J. M., & Krcmar, H. (2010). Does Collaboration among Participants Lead to Better Ideas in IT-based Idea Competitions? An Empirical Investigation. In *Proceedings of the Annual Hawaii International Conference on System Sciences* (pp. 1–10). https://doi.org/10.1109/HICSS.2010.157

Chernev, A. (2004). Goal-Attribute Compatibility in Consumer Choice. *Journal of Consumer Psychology*, *14*(1–2), 141–150. https://doi.org/10.1207/s15327663jcp1401

Dean, D. L., Hender, J. M., Rodgers, T. L., & Santanen, E. L. (2006). Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation. *Journal of the Association for Information Systems*, *7*(10), 646–698. https://doi.org/Article

Dörnyei, K. R., Krystallis, A., & Chrysochou, P. (2017). The impact of product assortment size and attribute quantity on information searches. *Journal of Consumer Marketing*, *34*(3), 191–201. https://doi.org/10.1108/JCM-10-2015-1594

Freitas, A. L., Higgins, E. T., Freitas, A. L., & Higgins, E. T. (2002). ENJOYING GOAL-DIRECTED ACTION : The Role of Regulatory Fit. *Psychological Science*, *13*(1), 1–6. https://doi.org/https://doi.org/10.1111%2F1467-9280.00401

Friedman, R. S., & Forster, J. (2001). The Effects of Promotion and Prevention Cues on Creativity. *Journal of Personality and Social Psychology*, *81*(6), 1001–1013. https://doi.org/1O.1O37//OO22-3514.81.6.1O01

Girotra, K., Terwiesch, C., & Ulrich, K. T. (2010). Idea Generation and the Quality of the Best Idea. *Management Science*, *56*(4), 591–605. https://doi.org/10.2139/ssrn.1082392

Glaholt, M. G., & Reingold, E. M. (2011). Eye Movement Monitoring as a Process Tracing Methodology in Decision Making Research. *Journal of Neuroscience, Psychology, and Economics*, *4*(2), 125–146. https://doi.org/10.1037/a0020692

Glöckner, A., & Herbold, A. K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, *24*(1), 71–98. https://doi.org/10.1002/bdm.684

Görzen, T., & Kundisch, D. (2017). When in Doubt Follow the Crowd: How Idea Quality Moderates the Effect of an Anchor on Idea Evaluation. *Thirty Eighth International Conference on Information Systems (ICIS)*, 1–20.

Gross, D. P. (2017). Performance feedback in competitive product development. *The RAND Journal of*

Economics, *48*(2), 438–466. https://doi.org/10.1111/1756-2171.12182

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis* (Seventh Ed). Harlow: Pearson Education Limited. https://doi.org/10.1038/259433b0

Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, *1*(1), 77–89. https://doi.org/10.1080/19312450709336664

Hensher, D. A. (2006). How do respondents process stated choice experiments? Attribute consideration under varying information load. *Journal of Applied Econometrics*, *21*(6), 861–878. https://doi.org/10.1002/jae.877

Higgins, E. T., Roney, C. J. R., Crowe, E., & Hymes, C. (1994). Ideal Versus Ought Predilections for Approach and Avoidance : Distinct Self-Regulatory Systems. *Journal of Personality and Social Psychology, 66*(2), 276–286. https://doi.org/http://dx.doi.org/10.1037/0022-3514.66.2.276

Hoornaert, S., Ballings, M., Malthouse, E. C., & Van den Poel, D. (2017). Identifying New Product Ideas: Waiting for the Wisdom of the Crowd or Screening Ideas in Real Time. *Journal of Product Innovation Management*, *34*(5), 580–597. https://doi.org/10.1111/jpim.12396

Jacoby, J. (1984). Perspectives on Information Overload. *Journal of Consumer Research*, *10*(4), 432–435.

Johnson, E. J., & Payne, J. W. (1985). Effort and Accuracy in Choice. *Management Science*, *31*(4), 395–414.

Kerstholt, J. (1992). Information search and choice accuracy as a function of task complexity and task structure. *Acta Psychologica*, *80*(1–3), 185–197. https://doi.org/10.1016/0001-6918(92)90046-G

Klein, N. M., & Yadav, M. S. (1989). Context Effects on Effort and Accuracy in Choice: An Enquiry into Adaptive Decision Making. *Journal of Consumer Research*, *15*(4), 411–421. https://doi.org/10.1086/209181

Kornish, L. J., & Ulrich, K. T. (2014). The Importance of the Raw Idea in Innovation: Testing the Sow's Ear Hypothesis. *Journal of Marketing Research*, *51*(1), 14–26. https://doi.org/10.1509/jmr.12.0401

Lohse, G. L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decision Processes*, *68*(1), 28–43. https://doi.org/10.1006/obhd.1996.0087

Magnusson, P. R., Netz, J., & Wästlund, E. (2014). Exploring holistic intuitive idea screening in the light of formal criteria. *Technovation*, *34*(5–6), 315–326. https://doi.org/10.1016/j.technovation.2014.03.003

Magnusson, P. R., Wästlund, E., & Netz, J. (2016). Exploring Users' Appropriateness as a Proxy for Experts When Screening New Product/Service Ideas. *Journal of Product Innovation Management*, *33*(1), 4–18. https://doi.org/10.1111/jpim.12251

Malhotra, N. K. (1982). Information Load and Consumer Decision Making. *Journal of Consumer Research*, *8*(4), 419. https://doi.org/10.1086/208882

Miller, G. A. (1956). The Magical Number Seven Plus or Minus Two: Some limits on Our Capacity for Processing Information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Mirsch, T., Lehrer, C., & Jung, R. (2017). Digital Nudging: Altering User Behavior in Digital Environments. *University of St. Gallen, Institute of Information Management*, *2*(12), 634–648.

Olsen, A. (2012). *The Tobii I-VT Fixation Filter: Algorithm description*. *Tobii Technology*. https://doi.org/10.1016/j.canep.2012.06.004

Olshavsky. (1979). Task Complexity and Contingent Processing in Decision Making : A Replication and Extension. *Organizational Behavior and Human Performance*, *24*(3), 300–316.

Payne, J. W. (1976). Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis. *Organizational Behavior and Human Performance*, *16*(2), 366–

387. https://doi.org/10.1016/0030-5073(76)90022-2

Payne, J. W., Bettman, J. R., Coupey, E., & Johnson, E. J. (1992). A constructive process view of decision making: Multiple strategies in judgment and choice. *Acta Psychologica*, *80*(1–3), 107–141. https://doi.org/10.1016/0001-6918(92)90043-D

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making.pdf. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534–552.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.

Payne, J. W., & Braunstein, M. L. (1978). Risky choice: An examination of information acquisition behavior. *Memory & Cognition*, *6*(5), 554–561.

Riedl, R., Brandstätter, E., & Roithmayr, F. (2008). Identifying decision strategies: A process- and outcome-based classification method. *Behavior Research Methods*, *40*(3), 795–807. https://doi.org/10.3758/BRM.40.3.795

Russo, J. E., & Dosher, B. A. (1983). Strategies for Multiattribute Binary Choice.pdf. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 676–696.

Schram, A., & Sonnemans, J. (2011). How individuals choose health insurance: An experimental analysis. *European Economic Review*, *55*(6), 799–819. https://doi.org/10.1016/j.euroecorev.2011.01.001

Sicilia, M., & Ruiz, S. (2010). The effects of the amount of information on cognitive responses in online purchasing tasks. *Electronic Commerce Research and Applications*, *9*(2), 183–191. https://doi.org/10.1016/j.elerap.2009.03.004

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. Retrieved from http://www.jstor.org/stable/1884852

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations little.* London: Little Brown.

Svenson, O. L. A. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance*, *23*(1), 86–112. https://doi.org/10.1016/0030-5073(79)90048-5

Swain, M. R., & Haka, S. F. (2000). Effects of Information Load on Capital Budgeting Decisions. *Behavioral Research in Accounting*, *12*(2), 171–198.

Terwiesch, C., & Ulrich, K. (2010). *Innovation Tournaments: Creating and Selecting Exceptional Opportunities*. *Journal of Product Innovation Management* (Vol. 27). Boston, MA: Harvard Business Press. https://doi.org/10.1111/j.1540-5885.2010.00739_2.x

Timmermans, D. (1993). The impact of task complexity on information use in multi-attribute decision making. *Journal of Behavioral Decision Making*, *6*(2), 95–111. https://doi.org/10.1002/bdm.3960060203

Toubia, O., & Netzer, O. (2017). Idea Generation, Creativity, and Prototypicality. *Marketing Science*, *36*(1), 1–20. https://doi.org/10.1287/mksc.2016.0994

Weinmann, M., Schneider, C., & Brocke, J. vom. (2016). Digital Nudging. *Business & Information Systems Engineering*, *58*(6), 433–436. https://doi.org/10.1007/s12599-016-0453-1