5-15-2019

THE FEWER, THE BETTER? EFFECTS OF DECOMPOSITION OF INFORMATION LOAD ON THE DECISION MAKING PROCESS AND OUTCOME IN IDEA SELECTION

Renata Santiago Walser
University of Innsbruck, renata.santiago-walser@uibk.ac.at

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

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

THE FEWER, THE BETTER? EFFECTS OF DECOMPOSITION OF INFORMATION LOAD ON THE DECISION MAKING PROCESS AND OUTCOME IN IDEA SELECTION

Research paper

Santiago Walser, Renata, University of Innsbruck, Innsbruck, Austria, renata.santiago-walser@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 are a growing trend among organizations that wish to harness the wisdom of crowds to achieve competitive advantage. Selecting the most promising ideas constitutes a challenge, as such contests generate hundreds or even thousands of ideas. In this context, it is increasingly important to use IT tools to support raters in the convergence process. Thus, it becomes essential to understand the decision processes associated with this task to develop platforms, which will nudge raters towards improved choice accuracy. Considering this goal, we conducted an online experiment in which 190 participants eliminated the least promising ideas in presentation modes with either high (2 ideas/screen) or low (30 ideas/screen) decomposition of information load. We found that higher decomposition of information load leads raters to acquire more information on ideas and exert more judgements. In turn, more judgements to eliminate ideas improved choice accuracy. Our findings add to the growing academic knowledge base on idea selection processes and how IT platforms can be designed to ensure successful convergence processes.

Keywords: Choice accuracy, Idea presentation, Innovation contests, Judgement

1 Introduction

Organizations show an increasing tendency of counting on the wisdom of crowds to generate innovative solutions using online innovation contests (Armisen and Majchrzak, 2015). Such contests usually encompass an idea generation phase and a subsequent selection process (Nagar, Boer and Garcia, 2016). A crowd often submits a substantial amount of ideas during generation. The selection phase involves the challenging task of narrowing the number of submissions from hundreds or thousands down to the few most promising ones, which constitutes a convergence process (Fu et al., 2017). But what makes a successful convergence process? The ability to select the best, high quality ideas into a converged set is an important quality determinant for the success of a convergence process (Girotra, Terwiesch and Ulrich, 2010) besides satisfaction, shared understanding, or convergence time (Davis, de Vreede and Briggs, 2007; Seeber, de Vreede, Maier and Weber, 2017). Assessing idea quality is a difficult endeavor. Therefore, researchers often compare the selection choices of raters with selection choices of experts. When selection choices of both types of raters are in agreement, choice accuracy is high and hence one could deduce the quality of the selected idea set. To maximize such choice outcomes, organizations must overcome the challenges associated with convergence. One key challenge in convergence is the high cognitive load required from raters to select the most promising ideas (Kolfschoten and Brazier, 2013) as they need to read, evaluate and decide on the merit of ideas. There exists a general assumption that more ideas are associated with higher cognitive load. However, past research could not confirm this assumption as multiple influencing factors are believed to affect this relationship and hence require additional research (Chernev, Böckenholt and Goodman, 2015). One such a factor may be the number of ideas presented simultaneously to evaluators, which can either ease or complicate choices (Johnson et al., 2012). This suggests that while showing all ideas at once imposes a certain information load, presenting subsets of ideas in multiple rounds could decompose this information load and may ease the task resulting in improved outcomes. Given the potential effects of decomposition of information load on choice accuracy, it remains unclear *how* this feature affects information processing towards the selection of the best ideas. Past research established that decision makers engage in processes of information acquisition, evaluation, action, and processing feedback/learning (Einhorn and Hogarth, 1981) when choosing between alternatives (ideas). Hence, the effects of an idea presentation mode that aims at decomposing information load may be explained by investigating people's decision making processes in terms of information acquisition, evaluation, action, and processing feedback/learning. Such an understanding is essential for the development of platform features that should nudge human information processing (Thaler and Sunstein, 2009) to improve the quality of idea convergence. Therefore, our study addresses the following research question: "What decision making processes do decision makers show when converging on ideas given different idea presentation modes?".

This study aims to fill this research gap by analyzing how two idea presentation modes that differ in their decompositions of information load (DIL) affect elements of raters' decision making processes and outcomes. We conducted an online experiment in which 190 participants eliminated ideas having either 30 ideas displayed at once (low DIL) or 15 subsets of 2 ideas at a time (high DIL). Our findings show that a high DIL is associated with more information acquisition and judgement in the decision making process, and that more judgement had a positive effect on choice accuracy. The findings confirm the relevance of idea presentation as antecedent of the judgement process that influences choice accuracy.

The remainder of the paper is organized as follows: in section 2, we provide the theoretical conceptualizations involving cognitive load and idea presentation in innovation contests and develop the hypotheses. Section 3 describes the experiment and explain the variables being measured. After describing the results in section 4, we discuss them while providing the contributions for theory and practice and limitations in section 5. Finally, we conclude the paper in section 6.

2 Background and Hypotheses

To successfully identify high-quality ideas, organizations make use of IT tools to facilitate idea selection (Girotra et al., 2010). Nevertheless, as they often make unrealistic demands of the crowd in terms of time, expertise, and cognitive effort, such approaches tend to perform poorly (Dellermann, Lipusch and Li, 2018). Thus, organizations often fail to exploit the crowd's true potential due to inadequate designs of evaluation tasks (Riedl, Blohm, Leimeister and Krcmar, 2010) and limited understanding of how and why interventions affect convergence outcome quality (Seeber et al., 2017). Cognitive load is a challenge during idea selection (Kolfschoten and Brazier, 2013), which is why it is important for convergence platforms to consider the cognitive load that a convergence task might impose on raters. According to Cognitive Load Theory (Sweller, 1988), cognitive load represents the mental effort deployed in working memory, and it is divided into *intrinsic*, *extraneous* and *germane* cognitive load (Paas, Renkl and Sweller, 2004). *Intrinsic* load is the cognitive load imposed by the task itself, being represented by elements such as the amount of information to be processed and the familiarity of the rater with the task. *Germane* load refers to how raters process available information in short and long-term memory. *Extraneous* load constitutes the way information is presented to raters. When a rater's cognitive load exceeds the rater's memory capacity in any of these dimensions (Paas et al., 2004), cognitive overload sets in and the rater cannot perform the task of selecting ideas accurately (Blohm, Riedl, Leimeister and Krcmar, 2011; Fu et al., 2017).

Given that cognitive load theory is concerned with techniques which help managing the working memory load (Paas et al., 2004), there is great potential for its use in Information Systems research. Particularly, the design of platforms that help raters keeping a low extraneous load (Fu et al., 2017) while increasing germane load (Paas et al., 2004) has a lot of room for research. Therefore, idea presentation is an important aspect of a choice architecture to consider when developing convergence platforms that help keeping cognitive load at manageable levels and thus improving choice outcomes.

When it comes to idea presentation, defining how many alternatives to present at the same time to raters represents a key factor in determining how decisions are made (Tversky and Kahneman, 1986; Johnson et al., 2012). Having an excessively high decomposition of information load, with too few ideas presented simultaneously (Johnson et al., 2012), or an excessively low decomposition of information load, with too many ideas presented at once (Schwartz et al., 2002; Willemsen, Graus and Knijnenburg, 2016) can be detrimental to the idea selection process. It is not clear how these two different variations of decomposition of information load nudge raters' human decision making in more in-depth processing and which choice outcomes this will yield.

In order to analyze how the number of ideas presented simultaneously might affect the choice process, one must first understand how decision makers choose ideas. First, selecting the most promising ideas demands compromise, judgement, and risk (Oman, Tumer, Wood and Seepersad, 2013). In addition, Einhorn and Hogarth (1981) divide the decision making process into four subprocesses: information acquisition, evaluation, action, and feedback/learning. In information acquisition, the decision maker searches for and storages information in memory and in the external environment. During evaluation, search strategies, such as maximization of expected value, elimination-by-aspects or satisficing, are used to process the acquired information. Action represents the final choice, being associated with a greater commitment than evaluation. Finally, feedback involves learning through the decision experience. In their decision making, raters may tend to apply more compensatory or more non-compensatory decision making according to the context (Pilli and Mazzon, 2016). Compensatory decision making describes raters that apply utility weights or values to the complete set of attributes and select the alternative with the best utility (Johnson and Payne, 1985). In this case, raters make a rational choice. In contrast, non-compensatory decision making does not consider all available attribute information or trades off the benefit of one attribute against the deficit of another attribute (Payne, Bettman and Johnson, 1993).

As raters narrow down their alternatives, they choose what information to consider and what information to ignore (Johnson et al., 2012). The presentation of a few alternatives allows a more reasoned comparison that does not overwhelm the decision maker (Johnson and Payne, 1985). In such cases, more compensatory decision processes are applied, which means that decision makers will acquire information on more attributes (Johnson and Payne, 1985). In contrast, when decision makers are presented with many alternatives, they tend to sacrifice information about certain attributes to process more alternatives (Dörnyei, Krystallis and Chrysochou, 2017). Hence, they search for less information across alternatives (Einhorn and Hogarth, 1981) and thus acquire information from fewer attributes. In the context of a pairwise idea presentation mode, where decomposition of information load is high, evaluators should tend to search for more information that describes an idea. Therefore, we hypothesize:

H1: Higher decomposition of information load will lead to more information acquisition.

A more deliberate reasoning process implies more effort when making trade-offs between alternatives (Johnson and Payne, 1985), which results in a more profound evaluation or judgement process (Einhorn and Hogarth, 1981). When decision makers are certain about the merit of an alternative, they are also able to judge. Thus, seeing someone choose an alternative or revise a choice may reflect that a decision maker arrived at a judgement (Einhorn and Hogarth, 1981). In contrast, when the evaluation process is cognitively too demanding and cognitive overload occurs, decision makers tend to defer their choice (Chernev et al., 2015). More judgement is expected when fewer ideas are presented at once as raters will be less likely to experience cognitive overload. This should become evident in their click behavior as we expect raters to perform the actions of "eliminating" and "keeping" ideas more often. Therefore, we hypothesize:

H2: Higher decomposition of information load will lead to more action of users to H2a "eliminate" and H2b "keep" ideas.

Tversky and Kahneman (1986) describe choice as a maximization process, in which rational decisions, when successfully achieved, lead to better decisions. Therefore, decision makers who choose ideas from a small number of ideas per screen will be guided towards acquiring more information about the available alternatives (Johnson and Payne, 1985). As a result, they will get closer to the maximizing goal, which should yield higher accuracy when the number of alternatives is not too excessive (Cui, Kumar Pm and Gonçalves, 2019). Thus, we hypothesize:

H3: More information acquisition will lead to higher choice accuracy.

The same applies for raters' actions. Raters that take action based on their evaluation processes are associated with more judgement, which, in turn, is linked with improved choice accuracy (Einhorn and Hogarth, 1981). As decision makers engage in more judgement, they are more certain of their decisions and, consequently, more likely to take action regarding the choice at hand (Lepora and Pezzulo, 2015). Therefore, we hypothesize:

H4: More action of users to H4a "eliminate" and H4b "keep" ideas will lead to higher choice accuracy.

As stated by Einhorn and Hogarth (1981), the decision making subprocesses are not independent from one another. When decision makers acquire information on attributes, they contrast this information with their existing mental schemas and if necessary extend their mental schema by storing newly acquired information (Van Merriënboer and Sweller, 2010). In addition, by decomposing the convergence task into smaller subtasks, the construction of mental schemas should be facilitated as decision makers feel less cognitively overloaded and are thus able to evaluate the ideas presented to them more efficiently. As more acquired information allows to build more comprehensive mental schemas, decision makers should feel certainty about their judgements, resulting in more choices to keep or eliminate ideas. Thus, it is expected that as information acquisition increases also the evaluation actions increase. Therefore, we hypothesize:

H5: Higher information acquisition will lead to more action of users to H5a "eliminate" and H5b "keep" ideas.

The introduced hypotheses are depicted in Figure 1 as a research model with several control variables that were found to affect the decision outcome in past research.

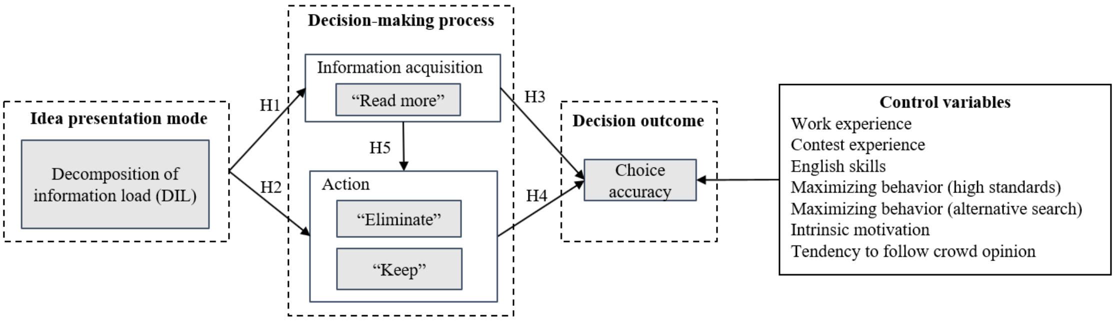

Figure 1. Research model

3 Method

To test the proposed research model, we conducted a between-subject online experiment with 190 participants. The experiment consisted of an idea selection task, in which participants reduced a set of 30 ideas by eliminating those they did not deem worthy of further consideration. In the following, we describe the experimental design in more detail.

3.1 Treatment variable: decomposition of information load

The decomposition of information load (DIL) represented how much information was displayed simultaneously to participants. It was operationalized with a distinct number of ideas per screen. The low DIL treatment displayed all 30 ideas at once. In the high DIL treatment, participants saw two ideas per screen and moved through 15 screens. This aimed to prompt them towards a more compensatory processing of ideas. In both treatments, the 30 ideas were presented in random order to avoid a position bias (Blohm et al., 2011). Figures 2 and 3 depict both treatments.

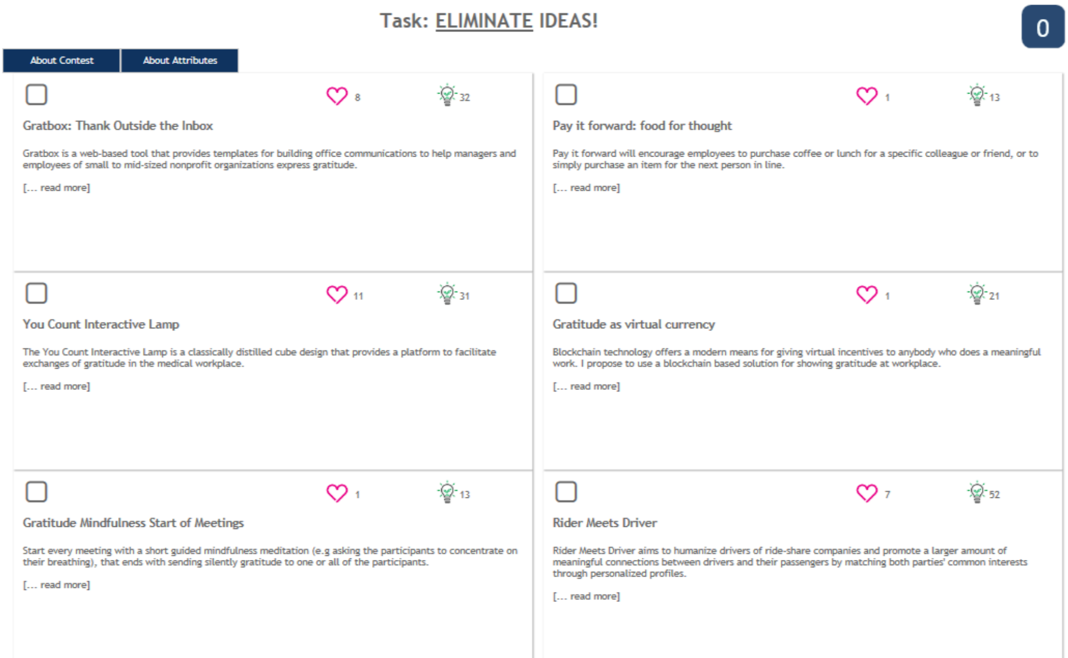

Figure 2. Low DIL treatment (30 ideas/screen)

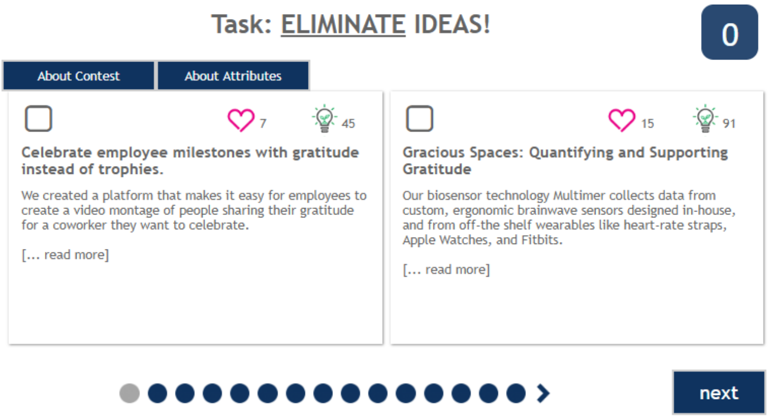

Figure 3. High DIL treatment (2 ideas/screen x 15 screens)

3.2 Subjects and sampling

We invited Business Administration students enrolled in the course "Introduction to Information Systems" at a European University to participate in the experiment. From the total of 240 students enrolled, 190 voluntarily took part in the experiment. Those who participated received extra course credits as a reward, provided that they answered the task attention check correctly. The online experiment was available to the participants for one week, and as soon as they accessed the platform, each participant was randomly assigned to one of the two treatment groups. We deemed students as appropriate participants because real-world idea convergence teams often rely on team members that have little domain knowledge of the contest domain (Merz et al., 2016).

3.3 Procedure, task and supporting idea selection platform

When participants accessed the platform, they were first redirected to the survey platform SoSci Survey, where they filled in an introductory survey containing the control variables. Subsequently, they were introduced to the platform and the task. In the introduction to the platform, we showed participants how an idea looked like and what information was available on every idea (see Figure 4). In addition, participants were informed that all information including the indices was collected from the original platform and that the contest was closed. Participants had no option to comment or applaud an idea or any other option that would change the information on ideas. We then specified the task with the prompt "You will see 30 ideas in total. In this phase, your task is to ELIMINATE ideas which do not seem promising for further consideration. How many ideas you eliminate is up to you. You can eliminate zero, one or multiple "bad" ideas from each set." and provided them with a page in which we outlined the contest goal through the prompt "how might we inspire experiences and expressions of gratitude in the workplace?" and did not specify selection criteria. Then, the experiment began, and participants went through the idea selection task in one of the two treatments of DIL. As mentioned in the prompt, participants were free to eliminate as many ideas as they wished. They could also change their minds and retain an idea they intended to select, as long as they did not confirm the elimination through the option "Finish". On every screen, the participants were able to visualize information about the contest and attributes from the idea cards through the options "About Contest" and "About Attributes".

While each participant went through the task, an activity log was saved in the database each time the "read more" or "eliminate" buttons were selected. A deselection of previously eliminated idea generated a "keep" log entry.

In order to assess the participants' honest commitment, the task included a task attention check in which participants had to select the option which described the task they just concluded. As soon as the participants eliminated ideas and confirmed the elimination, they moved on to the end survey to answer the task attention check. After that, the task was complete.

3.4 Idea set

We drew a stratified sample of 30 ideas from a real innovation contest on "expressions of gratitude in the workplace" which was hosted by the open innovation platform openIDEO in 2017. This contest topic was chosen, because business administration students are expected to be familiar with human resources issues and organizational settings. Thus, the contest topic should increase the ecological validity of the experiment (Pomerol and Adam, 2004). A small team of graduate and PhD students manually shortened each idea to about 120 words to control for idea length and reformulated the ideas to make sure that the short description of all ideas answered the questions a) "what is the idea about?" and b) "how does it work?". The convergence platform presented each idea with its title, description, and two feedback attributes. These attributes were the number of applauds (similar to likes) and the ideator score (ideator's past success). The values for both attributes were taken from the original platform. The purpose of adding these additional attributes to each idea's description was to keep the idea selection setting as realistic as possible to convergence processes in practice.

3.5 Measures

Information acquisition. Information acquisition encompasses the processes of searching and storing information (Einhorn and Hogarth, 1981). We measured this concept by counting how often a user clicked the "read more" button, which was added to each idea card. When participants were presented with the ideas they could only see a short introductory sentence of each idea, with about 30 words, as well as the "number of loves" and "idea score" (described in section 3.4). In case the participants wished to read the entire idea description and hence acquire more information, they could select the "read more" button and were presented with the remaining text.

Action. As taking action constitutes the final choice of alternatives after evaluating them (Einhorn and Hogarth, 1981), we used the "eliminate" and "keep" features to represent the action element of the decision making process. We measured "eliminate" with the number of clicks on the "eliminate" checkbox to indicate ideas "not worthy of further consideration". We measured "keep" as the number of clicks on the checkbox for re-including ideas into the consideration set that the participant had previously eliminated.

Figure 4 depicts how "read more" and "eliminate" were presented.

Figure 4. Idea card template

Choice accuracy. Choice accuracy refers to the effectiveness of each participant in identifying the best ideas from the idea set (Riedl et al., 2010). To operationalize this concept, we had to establish a quality benchmark, or "Gold Standard". Following previous research (Riedl et al., 2010; Magnusson, Netz and Wästlund, 2014), we adopted the consensual assessment technique (Amabile, 1982) and asked four domain experts to rate the ideas. Each rater assessed each idea in terms of its novelty, feasibility, elaborateness, and relevance which are well-established criteria to measure creativity of ideas (Dean, Hender, Rodgers and Santanen, 2006). After evaluation, the aggregated ratings were used to create a ranking of idea quality, from which the top 30% were considered "good ideas" (Blohm et al., 2011). After establishing the "Gold Standard", the *accuracy* performance measure used in information retrieval systems (Baeza-Yates and Ribeiro-Neto, 2011) was calculated. In the present context, accuracy consists of the fraction of ideas correctly classified as good (true positives, TP) and bad (true negatives, TN) divided by all ideas (see equation (1)).

$$Choice\ accuracy\ = \frac{TP + TN}{total\ number\ of\ ideas} \qquad \text{Equation (1)}$$

Control variables. W*ork experience* and *contest experience* were provided to control for the familiarity of the participants with the topic and the task, which could influence their cognitive load management and consequently their performance. *English language proficiency* was reported to analyze whether the degree of knowledge in English (foreign language) affected participants' performance. According to Schwartz et al. (2002), individuals known as *maximizers* (who always want the best alternative and are willing to search more for it) tend to compare more alternatives and spend more time and effort in the decision process, when compared to *satisficers* (who are satisfied with "good enough"). Maximizing behavior has two subdimensions which have been found to relate differently with various outcomes (Nenkov et al., 2008). Hence, we drew on the MTS-7 scale from Dalal, Diab, Zhu and Hwang (2015) to measure the "*high standards*" subdimension and the MMS scale from Lai (2010) to measure the "*alternative search*" subdimension. To measure the degree of subjects' involvement, *intrinsic motivation* items were adapted from the multidimensional work motivation scale (Gagné et al., 2015). Finally, to check whether the number of applauds and idea scores influenced the participants' outcomes, the ratio *tendency to follow crowd opinion* was created, consisting of the sum of applauds from the ideas which were not eliminated divided by the number of ideas which were not eliminated (see equation (2)). Thus, the higher the tendency to follow the crowd opinion is, the more likely participants were to consider the crowd's opinion when eliminating or keeping ideas. The items used to measure all constructs are listed in Appendix A: Survey.

$$Tendency\ to\ follow\ crowd\ opinion\ = \frac{\Sigma\ number\ of\ applauds\ from\ remaining\ idea\ set}{number\ of\ ideas\ in\ the\ remaining\ set} \qquad \text{Equation (2)}$$

4 Results

The present section describes the sample, reliability and validity tests, as well as the hypotheses testing.

4.1 Sample

A total of 190 individuals participated in the online experiment. 27 entries were removed because the participants did not answer the task attention check correctly, resulting in a sample of 163 cases. Subsequently, we checked for univariate and multivariate outliers (Hair et al., 2010) by calculating standardized scores and Mahalanobis distance. The analyses identified three cases with standardized and D^2/df values, respectively above 4 or below -4, which were considered outliers (Hair et al., 2010). After carefully assessing each outlier, only one case was excluded, as random responses were identified (Zijlstra, van der Ark and Sijtsma, 2011). The final sample with 162 participants was used for the subsequent analyses. Please refer to Table 1 for the sample description.

Item	Categories	Frequency	Percent
Age	18-21	93	57.4
	22-25	58	35.8
	26-29	11	6.8
	Total	162	100
Work experience (months)	0- <40	150	92.6
	40- <80	10	6.2
	80-120	2	1.2
	Total	162	100
Contest experience	yes, once	11	6.8
	yes, multiple times	4	2.5
	no, never	147	90.7
	Total	162	100
English language skills	Beginner	4	2.5
	Intermediate	47	29
	Advanced	90	55.6
	Proficient	21	13
	Total	162	100

Table 1. Age, work experience, contest experience and English language proficiency

From the final sample, 84 participants eliminated ideas in the low DIL treatment (30 ideas/screen). On average, they spent 7 minutes and 51 seconds on the elimination task, eliminated 13 ideas and obtained an accuracy of 53.06%. The remaining 78 participants went through the high DIL treatment (2 ideas/screen), who spent on average 10 minutes and 54 seconds on the task, eliminated 16 ideas and achieved an accuracy of 56.15%.

4.2 Reliability and validity

Results from confirmatory factor analysis (CFA) in SmartPLS 3 allowed us to assess internal consistency, convergent validity and discriminant validity (Hair Jr., Hult, Ringle and Sarstedt, 2014). (see Table 2). We assessed reliability of all multi-item constructs through the examination of individual item outer loadings and their internal consistency with Cronbach's alpha (α > 0.7) and composite reliability (CR > 0.70) (Hair Jr. et al., 2014). Six items of the High Standards (HS) construct and one item of the Alternative Search (AS) construct had to be excluded to achieve the respective Cronbach's alpha values of 0.690 and 0.681, which are considered acceptable in exploratory research (Hair Jr. et al., 2014). One item of the Intrinsic Motivation construct was eliminated to reach the common threshold for Cronbach's Alpha. For convergence validity, the outer loadings and average variance extracted (AVE) were examined. Items with outer loadings lower than 0.4 were dropped (Hair Jr. et al., 2014). Concerning average variance extracted (AVE), all scores exceeded the recommended threshold of 0.5 (Fornell and Larcker, 1981), with the exception of the HS construct (AVE = 0.490). Regarding discriminant validity, the Fornell-Lacker criterion was applied, and the square root of the AVE was compared with correlations between constructs. As the square root values of the AVE were greater than the off-diagonal correlations, discriminant validity was met (Fornell and Larcker, 1981). The results provide support for internal consistency, convergent validity and discriminant validity of our constructs.

Item	Outer loading	CFA														
		Correlations												Cr α	CR	AVE
		ACC	Action (elim)	Action (keep)	AS	Cont. Exp	DIL	English skills	HS	Info Acqu.	Int. Mot	Tend. to follow crowd op.	Work Exp.			
Accuracy		1.000														
Action (eliminate)		0.347	1.000													
Action (keep)		-0.092	0.417	1.000												
Alternative Search construct (AS)		0.100	0.087	-0.040	0.700									0.681	0.718	0.490
MA04_01	0.413															
MA04_02	0.576															
MA04_04	0.983															
Contest experience		-0.080	0.000	0.011	0.032	1.000										
DIL		0.202	0.345	0.399	-0.043	-0.096	1.000									
English skills		0.141	0.342	0.016	0.091	-0.136	0.048	1.000								
High Standards construct (HS)		0.148	0.134	-0.058	0.173	-0.051	-0.005	0.158	0.759					0.690	0.797	0.577
MA02_01	0.892															
MA02_03	0.542															
MA02_04	0.800															
Information acquisition		0.225	0.349	0.087	0.114	0.036	0.167	0.056	0.094	1.000						
Intrinsic Motivation		0.099	0.149	0.025	0.057	0.061	-0.050	0.162	0.160	0.169	0.869			0.744	0.859	0.756
MO02_02	0.979															
MO02_03	0.744															
Tendency to follow crowd opinion		-0.036	0.127	-0.031	-0.030	0.062	-0.136	0.072	0.085	0.014	0.184	1.000				
Work experience		0.018	0.151	0.034	0.023	0.053	0.041	-0.047	0.036	0.066	0.008	-0.252	1.000			

Table 2. Reliability and validity analysis

4.3 Hypotheses testing

Structural equation modelling (SEM) was adopted for hypotheses testing, using the software SmartPLS 3 (Hair Jr. et al., 2014). To assess the path model, we used path coefficients (β), path significance (p-value) and coefficient of determination (R^2). Path significance was calculated using a bootstrap resampling method with 5000 iterations.

The results of the hypotheses testing are shown in Table 3 and Figure 5. The hypotheses concerning the action element were broken down into two, to distinguish "eliminate" from "keep". In Figure 5, the significant effects are displayed, together with the path coefficients for the respective paths.

Hypotheses		Path Coefficient (β)	t-values	Significance level (p)	Status
H1:	DIL -> Information acquisition	0.167	2.160	p = .030	supported
H2a:	DIL -> Action (eliminate)	0.295	4.584	p < .001	supported
H2b:	DIL -> Action (keep)	0.396	6.152	p < .001	supported
H3:	Information acquisition -> Choice accuracy	0.095	1.282	not significant	not supported
H4a:	Action (eliminate) -> Choice accuracy	0.458	5.756	p < .001	supported
H4b:	Action (keep) -> Choice accuracy	-0.280	3.657	p < .001	not supported
H5a:	Information acquisition -> Action (eliminate)	0.299	4.937	p < .001	supported
H5b:	Information acquisition -> Action (keep)	0.021	0.293	not significant	not supported

Table 3. Hypotheses testing

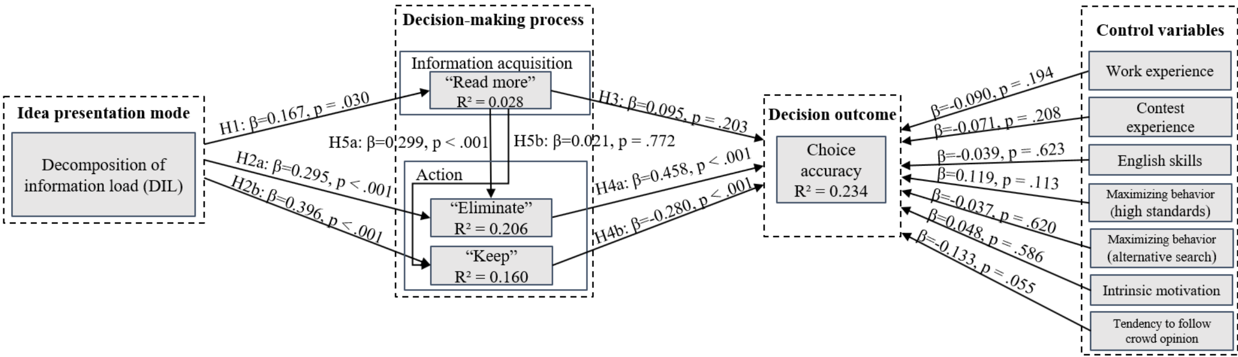

Figure 5. Results of hypotheses testing

As shown in Figure 5, the overall model explains 23.4 percent of variance in choice accuracy, which is considered weak (Hair Jr. et al., 2014). Moreover, none of the control variables had a significant effect on choice accuracy.

Hypothesis 1 suggested that a higher decomposition of information load (DIL) will be associated with more information acquisition. Our analysis revealed a significant relationship (β=0.167, p =.030) and, therefore, hypothesis 1 was supported.

Hypothesis 2 suggested that a higher DIL will be associated with more action. Significant relationships were found for 2a) action "eliminate" (β=0.295, p<.001) and 2b) action "keep" (β=0.396, p<.001), supporting hypotheses 2.

Hypothesis 3 tested whether more information acquisition would lead to higher choice accuracy. Results showed no significant effect of information acquisition on choice accuracy (β=0.095, p =.203). Therefore, hypothesis 3 was not supported.

Hypothesis 4 proposed that more action would be associated with higher choice accuracy. Constrasting results were found, as the hypothesis 4a, "eliminate" action, had a *positive* significant effect on choice accuracy (β=0.458, p<.001), while the hypothesis 4b, "keep" action, had a negative significant effect on choice accuracy (β=-0.280, p<.001). Thus, while the significant relationship of H4a is in line with our theorizing, the significant relationship of H4b was in the opposite direction. Therefore, hypothesis 4a was supported, but hypothesis 4b was not.

Finally, hypothesis 5 suggested that a higher information acquisition would lead to more action. Information acquisition had a positive significant effect on action "eliminate" (β= 0.299, p<.001), supporting hypothesis 5a. However, information acquisition had no significant effect on action "keep" (β= 0.021, p > .05), which means that hypothesis 5b was not supported.

5 Discussion

The present paper addressed the research question of what decision making processes do decision makers show when converging on ideas given different idea presentation modes? The model results revealed that presenting fewer ideas on a screen leads to more information acquisition and evaluation actions, which increases choice accuracy. These findings offer two contributions:

First, we contribute that an idea presentation mode with a higher decomposition of information load "nudges" raters to deliberate more on and judge more ideas, which leads to improved choice accuracy. Our findings show that raters who eliminated ideas given a high decomposition of information load (2 ideas/15 screens) acquired more information and eliminated more ideas compared to raters who saw all ideas at once. This offers unique insights into raters' decision making processes and suggests that choice accuracy can be improved when decision makers have to evaluate a high amount of ideas, as long as the information load is decomposed. Moreover, it could be that raters in our high DIL treatment were not only nudged into more information acquisition and judgements but also developed portable preferences (Johnson et al., 2012). Portable preferences describe a phenomenon where decision makers learn their preferences regarding attribute values from repeated choices. Amir and Levav (2008) found that repeated choices from 2-alternative sets reduced uncertainty surrounding trade-off values and improved choice consistency. In contrast, context-specific preference describes a phenomenon where decision makers choose based on overall preference instead of learning to weigh attributes between alternatives. Because the process of trading off was not learned, context-specific preferences make the choice less consistent across choice sets (Amir and Levav, 2008). Besides the manifestation of decision making processes, the development of portable preferences could provide additional insights to explain why and how higher DIL affects choice accuracy even if the total choice set has the same amount of ideas.

Second, we also provide empirical evidence to understand what decision making processes play a role in idea selection. More specifically, by being the first to investigate the decision making process within the idea selection context, our findings offer explanations as to how the different subprocesses of decision making (information acquisition and action) affect choice accuracy. Concerning the *elimination actions*, participants from the high DIL treatment eliminated more ideas and achieved higher choice accuracy. The higher engagement in judgement possibly led to a decrease in uncertainty, which made the elimination task easier and contributed to an increase in the reduction rate. Against our theorizing, more *information acquisition* did not lead participants to higher choice accuracy. Such finding corroborates Jamieson and Hyland's (2006) argument that additional information does not necessarily inform better decisions and reduce uncertainty. Cui et al. (2019) also had similar findings, in the sense that information acquisition was not directly associated with choice accuracy, but nudged participants to apply more effort to the task (more actions of “elimination”), which, in turn, increased choice accuracy. This indirect route of the effect of information acquisition on choice accuracy is in line with our findings.

Furthermore, we found that “keep” actions, i.e. reversing one’s decision to eliminate an idea, were negatively associated with choice accuracy. The findings of Lepora and Pezzulo (2015) could provide a reasonable explanation for this outcome. They suggested that, although changes of mind usually lead to higher accuracy, waiting too long to review one’s choice (i.e. revising after checking the whole set) might penalize choice accuracy. Hence, checking the “keep” element to keep alternatives open possibly reflected choice deferral in our study as a result of information overload (Pilli and Mazzon, 2016): Raters that went back and changed their minds might have considered the screens as interconnected choice sets and hence compared ideas across screens. Given that evaluating a set of 30 ideas is considered to be a cognitively demanding task (Iyengar and Lepper, 2000), they might have experienced cognitive overload and further deferred their choice. Deferring a choice is the opposite of taking action (Einhorn and Hogarth, 1981) and it has been found to diminish choice accuracy (Pilli and Mazzon, 2016).

5.1 Implications for theory and practice

Findings of the present research provide insights into how different idea presentation modes influence decision making and choice accuracy for idea selection. Research could build on our findings, as we show that the presentation of ideas matters. The fewer ideas presented, the better is choice accuracy. Hence, research on choice accuracy (e.g. Riedl et al., 2010; Cui et al., 2019) should be aware that the number of simultaneously presented ideas influences this phenomenon of interest. According to our findings, it is also advisable to present fewer ideas when establishing a Gold Standard for idea quality, because raters will engage in more information acquisition and judgement behavior (“eliminate” & “keep” actions) indicating a deeper processing of information and elaborate decision making.

Moreover, our findings also suggest a duality in the judgment behaviors “eliminate” and “keep”. We found a positive association between the “eliminate” action and choice accuracy, but a negative association between its opposite, the “keep” action, and choice accuracy. More research is required to understand whether these opposite effects of judgment behaviors constitute a peculiarity of this study or context, or under what conditions revising decisions (keep) should be avoided, for example in case the choice set is too big and might induce information overload and choice deferral.

Finally, by measuring information acquisition and judgement actions with activity logs, we provide a way to quantify elements of the decision making process. This allows measuring decision making behavior in idea selection experiments more objectively, which could foster theory development. Other researchers could adopt our measures to test their applicability in the same or related choice settings.

Our findings also have implications for innovation contest hosts and choice architects that are interested in the sociotechnical design of IT-enabled convergence mechanisms. Choice architects should be aware that they can nudge evaluators towards a more deliberate decision making process by decomposing the information load; showing fewer ideas on a screen. This is an inexpensive feature easy to implement, which, however, could improve the quality of idea selection. As convergence is a process that demands a great deal of effort and resources, the development of tools to help users being more accurate when selecting ideas holds great potential for increasing efficiency in the convergence process.

5.2 Limitations and avenues for future research

Although our investigation using experimental methods yielded an improved understanding of decision making, some limitations should be considered. First, even though the consensual assessment technique

(Amabile, 1982) has been often used to develop a Gold Standard for idea quality, such standards vary greatly across domain experts and studies. Therefore, more research in this area is necessary for the establishment of guidelines towards a consistent process of how to conduct idea quality measurements.

In addition, the small idea set of 30 ideas with reduced descriptions in the experiment might distance the choice task from a real-world convergence scenario. It was a necessary measure to ensure an amount of information load that would not surely cause choice paralysis (Álvarez, Rey and Sanchis, 2014), which would hinder choice accuracy. Future research could investigate whether present findings still hold in settings with more ideas, longer idea descriptions, or additional idea feedback attributes.

Furthermore, we did not measure the decision making process feedback/learning as participants did not receive outcome feedback during or at the end of the task. Hence, effects on choice accuracy might change when this process is considered. Future studies could assess which decision strategies raters use when they are presented with feedback and hence learn from their outcomes, which could influence choice accuracy.

Finally, although there was no significant effect of the “tendency to follow crowd opinion” on choice accuracy, it might be that this ratio affected the decision makers’ choice depending on the decomposition of information load of the task. Future research could analyze whether the presence of attributes such as “number of applauds” and “ideator score” had different effects on information acquisition and judgement actions for decision makers who are submitted to higher and lower decomposition of information load.

6 Conclusion

Overall, the present research shed light onto how to design idea convergence platforms in terms of decomposition of information load to increase the judgement process and consequently choice accuracy. Our findings add to the growing academic knowledge base on idea selection processes and how IT platforms can be designed to ensure a successful convergence process.

Appendices

Appendix A: Survey

Construct	Adapted items	Source
Personal information	What is your age?	own
	How many total months of work experience in a company do you have? (if no work experience, write zero)"	own
	Have you ever participated in an innovation contest? - yes, only once - yes, multiple times - no, never	own
	How would you rate your English proficiency? - beginner - intermediate - advanced - proficient	own
High standards	1. In studying or working, I always set the highest targets. 2. I don’t like having to settle for good enough. 3. I am a maximizer. 4. No matter what I do, I have the highest standards for myself. 5. I will wait for the best option, no matter how long it takes. 6. I never settle for second best. 7. I never settle. 8. No matter what it takes, I always try to choose the best.	Dalal et al. (2015)

	9. No matter how satisfied I am with my job, it is only right for me to be on the lookout for better opportunities.	
Alternative search	1. Whenever I'm faced with a choice, I try to imagine what all the other possibilities are, even ones that aren't present at the moment. 2. My decisions are well thought through. 3. I am uncomfortable making decisions before I know all of my options. 4. Before making a choice, I consider many alternatives thoroughly.	Lai et al. (2010)
Intrinsic motivation	Why did you put efforts or no efforts into eliminating/selecting ideas? Please indicate how much you agree or disagree with these statements. - Because the task was interesting. - Because the task was exciting. - I did little because I didn't think this task is worth putting efforts into. (reverse coded)	Adapted from Gagné et al. (2014)
Task attention check (Post-survey)	What was the task you just concluded? - eliminated bad ideas - selected good ideas - none of the above	own

References

Álvarez, F., J. M. Rey and R. G. Sanchis. (2014). "Choice overload, satisficing behavior, and price distribution in a time allocation model." *Abstract and Applied Analysis*, *2014*, 1–9.

Amabile, T. M. (1982). "Social Psychology of Creativity : A Consensual Assessment Technique." *Journal of Personality and Social Psychology*, *43*(5), 997–1013.

Amir, O. and J. Levav. (2008). "Choice construction versus preference construction: the instability of preferences learned in context." *Journal of Marketing Research*, *45*(2), 145–158.

Armisen, A. and A. Majchrzak. (2015). "Tapping the innovative business potential of innovation contests." *Business Horizons*, *58*(4), 389–399.

Baeza-Yates, R. and B. Ribeiro-Neto. (2011). *Modern information retrieval: the concepts and technology behind search* (2nd ed.). New Jersey, USA: Addi-son-Wesley Professional.

Blohm, I., C. Riedl, J. M. Leimeister and H. Krcmar. (2011). "Idea evaluation mechanisms for collective intelligence in open innovation communities: Do traders outperform raters?" In: *Proceedings of 32nd International Conference on Information Systems 2011*.

Chernev, A., U. Böckenholt and J. Goodman. (2015). "Choice overload: A conceptual review and meta-analysis." *Journal of Consumer Psychology*, *25*(2), 333–358.

Cui, Z., S. Kumar Pm and D. Gonçalves. (2019). "Scoring vs. Ranking: An Experimental Study of Idea Evaluation Processes." *Production and Operations Management*, *28*(1), 176–188.

Dalal, D. K., D. L. Diab, X. S. Zhu and T. Hwang. (2015). "Understanding the construct of maximizing tendency: A theoretical and epirical evaluation." *Journal of Behavioral Decision Making*, *28*(5), 437–450.

Davis, A. J., G.-J. de Vreede and R. O. Briggs. (2007). "Designing thinkLets for convergence." In: *AMCIS 2007 Proceedings* (p. 358).

Dean, D. L., J. M. Hender, T. L. Rodgers and E. L. Santanen. (2006). "Identifying Quality , Novel , and Creative Ideas : Constructs and Scales for Idea Evaluation." *Journal of the Association for Information Systems*, *7*(10), 646–699.

Dellermann, D., N. Lipusch and M. Li. (2018). "Combining Humans and Machine Learning: A Novel Approach for Evaluating Crowdsourcing Contributions in Idea Contests." In: *Multikonferenz Wirtschaftsinformatik (MKWI)*. Lüneburg, Germany.

Dörnyei, K. R., A. Krystallis and P. Chrysochou. (2017). "The impact of product assortment size and attribute quantity on information searches." *Journal of Consumer Marketing*, *34*(3), 191–201.

Einhorn, H. J. and R. M. Hogarth. (1981). "Behavioral Decision Theory: Processes of Judgment and Choice." *Annual Review of Psychology*, *32*(1), 53–88.

Fornell, C. and D. F. Larcker. (1981). "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error." *Journal of Marketing Research, 18*(1), 39–50.

Fu, S., G.-J. de Vreede, X. Cheng, I. Seeber, R. Maier and B. Weber. (2017). "Convergence of Crowdsourcing Ideas : A Cognitive Load Perspective." In: *38th International Conference on Information Systems: Transforming Society with Digital Innovation*. Association for Information Systems.

Gagné, M., J. Forest, M. Vansteenkiste, L. Crevier-Braud, A. van den Broeck, A. K. Aspeli, … C. Westbye. (2015). "The Multidimensional Work Motivation Scale: Validation evidence in seven languages and nine countries." *European Journal of Work and Organizational Psychology*, *24*(2), 178–196.

Girotra, K., C. Terwiesch and K. T. Ulrich. (2010). "Idea generation and the quality of the best idea." *Management Science*, *56*(4), 591–605.

Hair, J. F., W. C. W. C. Black, B. J. Babin, R. E. Anderson, B. J. Babin and W. C. W. C. Black. (2010). *Multivariate data analysis: A global perspective* (Vol. 7). Pearson Education.

Hair Jr., J., G. T. M. Hult, C. M. Ringle and M. Sarstedt. (2014). *A primer on partial least squares structural equations modeling (PLS-SEM)*. SAGE.

Iyengar, S. S. and M. R. Lepper. (2000). "When Choice is Demotivationg: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology*, *79*(6), 995–1006.

Jamieson, K. and P. Hyland. (2006). "Good intuition or fear and uncertainty: The effects of bias on information systems selection decisions." *Informing Science*, *9*, 49–69.

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

Johnson, E. J., S. B. Shu, B. G. C. Dellaert, C. Fox, D. G. Goldstein, G. Häubl, … E. U. Weber. (2012). "Beyond nudges: Tools of a choice architecture." *Marketing Letters*, *23*(2), 487–504.

Kolfschoten, G. L. and F. M. T. Brazier. (2013). "Cognitive Load in Collaboration: Convergence." *Group Decision and Negotiation*, *22*(5), 975–996.

Lai, L. (2010). "Maximizing without difficulty: A modified maximizing scale and its correlates." *Judgment and Decision Making*, *5*(3), 164–175.

Lepora, N. F. and G. Pezzulo. (2015). "Embodied Choice: How Action Influences Perceptual Decision Making." *PLOS Computational Biology*, *11*(4), 1–22.

Magnusson, P. R., J. Netz and E. Wästlund. (2014). "Exploring holistic intuitive idea screening in the light of formal criteria." *Technovation*, *34*(5–6), 315–326.

Merz, A. B., I. Seeber, R. Maier, A. Richter, R. Schimpf, J. Füller and G. Schwabe. (2016). "Exploring the Effects of Contest Mechanisms on Idea Shortlisting in an Open Idea Competition." In: *Thirty Seventh International Conference on Information Systems* (pp. 1–18). Dublin.

Nagar, Y., P. de Boer and A. C. B. Garcia. (2016). "Accelerating the Review of Complex Intellectual Artifacts in Crowdsourced Innovation Challenges." In: *Thirty Seventh International Conference on Information Systems* (pp. 1–17).

Nenkov, G. Y., M. Morrin, A. Ward, B. Schwartz and J. Hulland. (2008). "A short form of the Maximization Scale: Factor structure, reliability and validity studies." *Judgment and Decision Making*, *3*(5), 371–388.

Oman, S. K., I. Y. Tumer, K. Wood and C. Seepersad. (2013). "A comparison of creativity and innovation metrics and sample validation through in-class design projects." *Research in Engineering Design*, *24*(1), 65–92.

Paas, F., A. Renkl and J. Sweller. (2004). "Cognitive Load Theory : Instructional Implications of the Interaction between Information Structures and Cognitive Architecture." *Learning and Instruction, 32*(1), 1–8.

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

Pilli, L. E. and J. A. Mazzon. (2016). "Information overload, choice deferral, and moderating role of need for cognition: Empirical evidence." *Revista de Administração, 51*(1), 36–55.

Pomerol, J. and F. Adam. (2004). "Practical Decision Making – From the Legacy of Herbert Simon to Decision Support Systems." In: *Actes de la Conférence Internationale IFIP TC8/WG8* (pp. 647–657).

Riedl, C., I. Blohm, J. M. Leimeister and H. Krcmar. (2010). "Rating scales for collective intelligence in innovation communities: Why quick and easy decision making does not get it right." In: *Thirty First International Conference on Information Systems*. St. Louis.

Schwartz, B., A. Ward, S. Lyubomirsky, J. Monterosso, K. White and D. R. Lehman. (2002). "Maximizing versus satisficing: Happiness is a matter of choice." *Journal of Personality and Social Psychology*, *83*(5), 1178–1197.

Seeber, I., G. J. de Vreede, R. Maier and B. Weber. (2017). "Beyond Brainstorming: Exploring Convergence in Teams." *Journal of Management Information Systems, 34*(4), 939–969.

Sweller, J. (1988). "Cognitive load during problem solving: Effects on learning." *Cognitive Science, 12*(2), 257–285.

Thaler, R. H. and C. R. Sunstein. (2009). *Nudge : improving decisions about health, wealth and happiness*. Penguin Books.

Tversky, Am. and D. Kahneman. (1986). "Rational Choice and the Framing of Decisions." *The Journal of Business*, *59*(4), 251–278.

Van Merriënboer, J. J. G. and J. Sweller. (2010). "Cognitive load theory in health professional education: Design principles and strategies." *Medical Education*, *44*(1), 85–93.

Willemsen, M. C., M. P. Graus and B. P. Knijnenburg. (2016). "Understanding the role of latent feature diversification on choice difficulty and satisfaction." *User Modelling and User-Adapted Interaction*, *26*(4), 347–389.

Zijlstra, W. P., L. A. van der Ark and K. Sijtsma. (2011). "Outliers in Questionnaire Data: Can They Be Detected and Should They Be Removed?" *Journal of Educational and Behavioral Statistics*, *36*(2), 186–212.