



Evaluation of Early-Stage Design Concepts via Online Discussion: A Case Study

Chiara Leonardi, Massimo Zancanaro
Fondazione Bruno Kessler
(FBK)
Trento, Italy

Anu Seisto
VTT, Technical Research Centre
(VTT)
Espoo, Finland

Nicola Doppio
Hub Innovazione Trentino
(HIT)
Trento, Italy

Aino Mensonen
VTT, Users and Media
(VTT)
Espoo, Finland

Abstract

User involvement through an online platform may have several benefits for evaluating early-stage design ideas. In this paper, we present a case study where we supported two companies with gathering customers' early feedback on service concepts. Our findings suggest that a scenario-based approach coupled with an explicit rewarding mechanism based on actual participation may prompt vivid discussions, reduce dropouts and "lurkers" as well as provide useful insight into the evaluation of services and elicit new ideas.

Published 29 June 2018

Correspondence should be addressed to Massimo Zancanaro, Fondazione Bruno Kessler (FBK), Trento, Italy. Email: zancana@fbk.eu

DigitCult, Scientific Journal on Digital Cultures is an academic journal of international scope, peer-reviewed and open access, aiming to value international research and to present current debate on digital culture, technological innovation and social change. ISSN: 2531-5994. URL: <http://www.digitcult.it>

Copyright rests with the authors. This work is released under a Creative Commons Attribution (IT) Licence, version 3.0. For details please see <http://creativecommons.org/licenses/by/3.0/it/>



Introduction

Early evaluation of design concepts is crucial to understand as soon as possible customers' reactions toward new products and services. Traditionally, this is done through the so-called "concept testing" which consists in asking a large group of potential customers to rate short descriptions of ideas on one or more ranking scales (Moore 1982). This approach has some limitations and it is often partially replaced by in-depth interviews and focus groups with smaller groups of customers that may provide more informative feedback (Iuso 1975; Cox et al. 1976). Interest in online platforms for discussion has recently increased as they may offer the opportunity to conduct large focus groups beyond the boundaries of time and physical location (Janzik et al. 2011). In particular, in the context of the open innovation approach, companies strive to identify lead users and early adopters to include them in the product development process (von Hippel, 1986). In the open innovation approach, there are at least two ways in which customers can be involved in online discussions. The first one is as innovation communities, formed by lead users with strong innovative skills. The second is as brand communities, made up by people who have a strong connection with the brand and whose loyalty can be maintained for example through a Facebook fan page (Zeng 2014). With respect to face-to-face focus groups, online group discussions tend to elicit more comments, although shorter and simpler (Schneider et al. 2012). Furthermore, they facilitate access to participants who are difficult to recruit and might even reduce their inhibition to participate (Murgado-Armenteros 2012). Yet, there might be challenges and drawbacks too. Online contributions of participants may be less complex and rich than in face-to-face discussions and feedback collected may be less useful (Cowley et al. 2011). Participants may be less motivated to detail their opinions and the reasons behind them, as well as less motivated to build on others' participants' contributions (Følstad 2016).

In this paper, we discuss an experience with an online discussion platform used for three weeks within an industrial innovation project to the end of involving potential customers in structured discussions on digital services in the retail sector. The goal of the experience was to collect early-stage feedback on offered services as well as eliciting novel ideas. Our approach differs from those presented in the literature because (i) we used an explicit reward mechanism based on actual participation and (ii) we structured the discussion by using design concepts scenarios and guiding questions.

This paper aims at contributing to the ongoing research on the use of online platforms for early design concept evaluation by investigating three research questions that we believe are only partially considered in the literature:

- *RQ1: does this approach foster rich collaborative discussions (rather than meager and mainly unrelated individual contributions)?*
- *RQ2: do participation and patterns of contribution of the community change over time?*
- *RQ3: is this approach suited to assess design concepts and to elicit new ideas?*

We followed a case study approach (Yin 2009) in order to gain in-depth understanding of the phenomenon and its context (Cavaye 1996). Specifically, we explored the possibility of capturing the 'reality' by observing the experience of participants and companies involved in a natural context rather than constructing a lab-based experience.

Our case study provides evidence that a rewarding mechanism and structured discussion may limit participants' drop-out rates and "lurkers" (participants who browse discussions without actively contributing to them; Amichai-Hamburger et al. 2014) while fostering effective interactions and conversations.

The paper is organized as follows: in the next section, we discuss some related works regarding the assessment of early-stage design concepts, and the engagement of people in online tasks. We then present the details on how we organized and run our study and provide a thorough analysis of the discussion dynamics as emerged from the online platform logs as well as a content analysis aimed at investigating the richness of the discussions. Finally, we discuss the lessons learned from this experience.

Related Works

This work contributes to and leverages from previous research on early-stage design concept evaluation and on studies that exploit and reflect on online and crowd approaches for user engagement. We also look at works comparing traditional face-to-face methods (such as focus groups) and online tools.

Traditional user-centered design and participatory design approaches (among others, Norman 1986; Beyer & Holtzblatt 1998; Bødker 2000) stress the importance of user-designer interaction which is often attained through face to face discussions (interviews, focus groups, workshops, etc.) to foster mutual understanding. However, there is growing interest in web-based methods that enable rich interaction and flexible user participation in different stages of the design process (Ma 2015).

A number of studies have investigated the strengths and limits of online platforms as a method for exploring people's attitudes, reactions toward products, etc. (see Malinen 2005 for a survey of research on online communities). Näkki (2012) suggests that online tools offer opportunities for involvement of geographically distributed actors and for long-term and continuous contributions. Researchers and designers can keep close contact with users, monitor the process and intervene directly when needed (Hagen 2011; Yndigegn 2010). Reyes (2012) also pointed out that adapting traditional methods for online environment is useful for collectively producing and evaluating design ideas and to guarantee the participation of a heterogeneous panel of participants. Online methods offer the possibility of conducting asynchronous research, i.e. the possibility of moderating discussion over long periods of time. Furthermore, online tools may offer the opportunity to reach remote, reluctant, or unconventional groups of participants, favor anonymity among participants, and facilitate the discussion about personal and sensitive issues, reducing the risk of inhibition (Nunamaker 1996).

On the other hand, the online approach presents a number of drawbacks: lack of user representativeness (as Internet users only participate) and participants' high dropout rates (Schneider 2002). Schneider (2002) also discusses two key differences between online and face-to-face focus groups that have an impact on the type and quality of conversation: media richness and social presence. Online tools often do not adequately support media richness, defined as the ability of a communication medium to foster immediate interaction and support people in expressing their ideas by using cues. Social presence - the degree to which a medium conveys the immediacy of face-to-face conversation - is also scarcely supported by online tools. These differences may have several consequences in the type of discussion they stimulate. In the online setting, participants may contribute with more comments because they do not receive negative cues nor are interrupted by another people. Contribution may be less complex than in face-to-face discussions and participants may be less motivated to detail their opinions and reasons behind them (Murgado-Armenteros 2012). In their study, Følstad et al. (2016) used an online social platform and found that an important limit was that participants did not build on each other's contributions to express their ideas.

Ma et al. (2015) found that guidance and structure is crucial for collecting insightful user data. They compared "need-finding" activities conducted with both face-to-face and online participants on similar budget constraints. They explored whether online videos may support the production of feedback on preliminary concepts storyboards and found that online communication conducted in parallel with face-to-face interaction is effective to quickly collect users' insights.

Some studies also investigated the motivations and rewards that support the collection of high-quality design feedback. Financial rewards seem to play an important role in supporting user feedback generation. Yen et al. (2016) explore differences in feedback received by participants exploiting different online environments and leveraging on different motivations: i) social networks (social crowd), ii) online markets such as Amazon's Mechanical Turk (financial crowd), and iii) web forum (enjoyment crowd). They measured feedback according to its quality, quantity and valence and found that the financial crowd produced longer and more positive feedback and provided more design suggestions. Greenberg et al. (2015) evaluated Critiki, a tool for gathering design critiques for crowdfunding project creators: by a controlled experiment with 450 crowd

workers, they demonstrated the efficacy of a small amount of money to collect high-quality feedback.

The Case Study

The case study discussed in this paper was part of a larger project commissioned by a service company and a large retail company with the aim to investigate new ICT-based services in the retail sector.

The companies identified three design ideas for services that they were interested to evaluate and refine before implementing them for the retail company's customers. All the envisioned services had in common the ultimate goal of using personal information already collected by the retail company to improve customer satisfaction and retention in the retail chain. In the end, two of them were implemented as pre-commercial prototypes as a result of the online discussion.

The experiment was organized as a 3-week online structured discussion among a relatively large group of (potential) customers of the retail chain. The aim of the discussion was to elicit positive and negative attitudes toward the design ideas. Table 1 summarizes the design ideas as proposed by the companies.

Table 1. The three design ideas proposed for assessment by the companies.

<p>DESIGN IDEA 1: Valuing customers' shopping habits data</p> <p>Each retail company has a large number of data on customers' purchases, which are routinely used to optimize business decisions. This design idea aimed to exploit this data to offer new services to customers. Services can range from simple visualization of the products purchased in the store, to comparison of one's own shopping habits to other customers' habits, or provision of personalized suggestions for new products.</p>
<p>DESIGN IDEA 2: Exploiting a Personal Data Store</p> <p>Each retail company stores data about customers in proprietary databases. Although customers have the legal rights to inspect or cancel the data, which belong to them, the process is usually quite cumbersome.</p> <p>This design idea explored the possibility of employing a Persona Data Store, a secure digital web app available for each customer that would provide a private repository for personal information. The services that might be offered through Personal Data Stores range from controlling which data to make visible and which to hide to exploring the possibility of selling to third parties the right to use such data.</p>
<p>DESIGN IDEA 3: Gamifying the shopping experience</p> <p>The shopping experience, in particular grocery shopping, is usually boring and tedious.</p> <p>This design idea explored the possibility of embedding gaming experiences as part of the shopping experience. Examples of gamification are both a treasure hunt to find discounted products and a crowdsourcing activity to categorize interesting features of purchased products.</p>

Materials

In order to make the design ideas easier to communicate and encourage participants' reflection, we developed 3 to 4 short stories for each design idea.

The use of short stories, also called scenarios, has a long tradition in User-Centered Design (Carroll and Rosson 1992; Carrol 2000; Rosson and Carroll 2002). A scenario is a story that narrates the use of a future system (or service) from the user's point of view. Scenarios help shifting the focus of the design work from defining system operations to understanding how people will use a system (Rosson and Carroll 2002). We employed scenarios to describe typical situations of users engaged in the envisioned services and framed in a neutral perspective and we also make use of personas (Cooper 1994) to foster participants' identification and engagement. The stories were drafted by the research team, and then discussed and further elaborated with the companies.

The goal of the scenarios was to exemplify the design ideas and to engage the participants in a discussion aimed at exploring their attitudes and concerns toward the services. Furthermore,

the scenarios were expected to promote discussions about new opportunities offered by the type of technologies presented as well as to collect novel ideas.

Figure 1 shows a persona and one of the scenarios realized. In total, 11 scenarios were produced for the three design ideas. The scenarios were text-based short stories enriched with some simple visual representation of the service, usually as screenshots of a fictitious mobile app or web page (see appendix).




LAURA	SCENARIO: Know more about your spending habits
 <p>This week we present Laura (36yo). She works as a teacher and she has two kids. In this scenario, you will see how Laura will use a new mobile app to do her grocery shopping. She would like to better understand how she spends her money, so she can still save money without giving up quality.</p>	<p>Laura has just downloaded a new app for which she already has a loyalty card. With the help of the app, Laura may view her expenses.</p>  <p>Laura has the impression of having spent more than usual this month. Looking at the graph, she notices that she spent much more than the previous months.</p>  <p>Looking at the graph on the app, Laura immediately understands why she spent so much last month: she bought 20% more organic food...</p>

Figure 1. An example of a persona and a scenario published on the online platform (all the scenarios are reported in the Appendix).

In order to bring the discussions into focus, we added, for each scenario, a few questions aimed at exploring specific dimensions. In particular, some questions meant to elicit the reaction of participants on the proposed service (evaluation questions) or to trigger participants' reflection on alternative services (creative questions).

With evaluation questions, we aimed at investigating, for each design idea: i) perceived usefulness and attractiveness, ii) possible issues related to the use of personal data (privacy, ethics, trust). An example of an evaluation question is the following: "Would you trust a service like the one described that collects personal data on your purchases?"

With ideation questions, we aimed at encouraging participants to produce new ideas starting from the scenario proposed. An example of this type of question is: "Starting from the proposed scenario, can you envisage other ways to use your spending data to improve the way you choose grocery products?"

In total, 42 questions were included (3 to 5 for each scenario): 24 were evaluation-type questions and 18 were creativity-type ones. The full list of the probe questions is reported in the appendix.

Owela Online Innovation Space

The platform used for the online discussion was Owela (Friedrich 2013), an open innovation and co-design online platform.

The platform allows the active involvement of users in the innovation process from the evaluation of early ideas up to the management of feedback during piloting and actual use. Based on the *Wordpress* open source content management system, it is a constellation of plugins that can be flexibly integrated, depending on the specific features required in each project workspace. It allows administrators to post different types of contents (text, pictures and videos) as input for discussion. Focus group discussions can be organized either asynchronously using a "discussion forum" structure or synchronously as a chat. Other data collection methods, such as questionnaires and quick polls, can be added to the discussions.

For the study presented in this paper, we employed an asynchronous discussion structure. Administrators would regularly post scenarios and questions to bring the discussion into focus. Participants would post their comments as answers to the questions or as replies to other participants' comments. Participants could also express their attitude toward other participants' comments by expressing approval ("thumb up") or disapproval ("thumb down"). Since we

rewarded participants with financial incentives according to their effective participation, the platform displayed individual scores on each participant's profile page.



Figure 2. A screenshot of the online discussion space. At the top (1), the question posted by the administrator to trigger the discussion. Below, (2), three posts by participants. The last post, (3), is a user's reply to another participant's comment. Thumbs up and down given by other participants (4) are shown to the right of each post.

Recruitment of Participants

The participants were involved in the project using the Smart Crowds Territorial Lab a "Living Lab" facility based in Trento, Italy, and aimed at supporting research centers and companies in setting up and managing living lab projects (Leonardi et al. 2014).

The recruitment for the project was advertised on the Living Lab webpage and a call was sent to Living Lab long-term subscribers. Candidates were required to complete online application forms. The recruiting process lasted for one month. It included a referral policy option by which subscribers could invite a friend to submit their application, thus creating a snowball effect. Over 550 requests for participation were received.

Eventually, we had to limit participation to 140 people due to budget constraints. Participants were selected using a profile-based sampling plan aimed at ensuring representativeness of the target population (adults with familiarity with ICT and internet access at home, statistical data were taken from the Italian National Statistical Institute).

Rewarding Mechanism

There have been much debate in the literature about how monetary rewards may potentially benefit or harm performance in online communities (Bonner et al. 2000; Janzik and Raash 2011). According to the Self-Determination Theory (Ryan and Deci 2000), a distinction must be made between extrinsic and intrinsic motivation: the former is based on external incentives such as financial rewards while the latter derives from a state of inherent satisfaction. In a seminal work, Deci (1971) argues that financial incentives may reduce intrinsic motivation to perform some types

of activities. In discussing the motivation to participate in online communities of customers, Janzik and Raash (2011) suggest that extrinsic motivation by itself is not sufficient to explain innovative activities and that financial rewards may play a minor role in the extrinsic motivation of participants.

Indeed, Heyman and Ariely (2004) argue that there are two types of markets that determine relationships between effort and payment: one is the monetary market and the other one is the social market. For the latter, payments may be detrimental but not for the former. In our case, since we were not working with a brand community but with customers willing to participate in an online discussion instead, we assumed that our community was closer to a monetary market rather than a social one. Therefore, we deemed necessary to implement an extrinsic reward mechanism. Indeed, the type of community we are dealing with is more similar to crowd workers than to a brand community and there is some evidence that for this kind of communities, financial incentives are very important (among others, Behrend and colleagues 2011).

For this type of setting, the more effective scheme was based on quota rather than on piece-rate or fixed-rate (Bonner et al. 2000; Mason and Watts 2009). We therefore designed a quota scheme where the participants received a weekly reward when they reached a certain amount of contributions and the reward did not increase with further activity.

A scoring procedure was integrated into *Owela* that computed participants' contribution as follows: 1 point for each "thumb up" (like) or "thumb down" (dislike) on the comment of another participant; 3 points for each comment on a scenario or reply to the comment of another participant. Each week, participants were required to score a minimum of 10 points to receive their compensation (a 20€ Amazon voucher). The updated score was accessible to each participant from their personal profile page at all times.

Procedure

The study lasted 3 weeks. Each week was dedicated to exploring one of the 3 concept ideas through several scenarios (as described in Table 2) by means of a number of probe questions. All the scenarios for each design idea (together with the associated questions) were published on three consecutive Mondays, and stayed active until the end of the study. An e-mail was sent to participants each Monday to notify the publication of a new set of scenarios (and the relating probe questions).

Table 2. Scenario presentation distribution in the three week period the (scenarios are reported in the appendix).

Week 1	Week 2	Week 3
Design idea: Valuing the data about shopping	Design idea: Using a personal data store	Design idea: Gamification of the shopping experience
Scenarios: myshop1, myshop2, myshop3, myshop4	Scenarios: pds1, pds2, pds3	Scenarios: myshop5, myshop6, myshop7, myshop8

The very first day of the study, 95 people logged into the system (68% of the total), and by the second day 91% of the participants had logged in (128 people). At the end of the first week, 139 participants had logged into the platform. The last one joined at the beginning of the third week. No participants dropped out.

Data Collection and Preparation

The data collected through the *Owela* platform consisted of all the comments posted in the period and the thumbs up or down received with each post. For each post, the timestamp, the author (we used authors' identifier to ensure privacy), and the relevant scenario and question were recorded. Furthermore, for each response-to-comment post, we also collected the identifier of the original post. Figure 3 shows an extract from the dataset.

In order to get a better understanding of the content discussed, all the comments were analyzed independently by two researchers of the team, using a Grounded Theory approach (Glaser and Strauss 1967).

Initially, researchers individually annotated the emerging topics using the so-called “open coding” approach. For each comment, researchers assigned one or more codes that summarized the main aspect discussed in that comment. For example, the comment “*The fact that you have a fidelity card means that you have voluntarily given your personal data to the supermarket [company]. I don’t think that there are additional risks [in using the service under discussion]. Of course, there is the possibility that the supermarket [company] send you unsolicited advertising about products you recently bought [...]*” contains some reference to potential privacy issues related to sharing personal information with the retail company, and it also mentions the possibility that the company will use this information for advertising: the appropriate codes could be “privacy”, “data sharing” and “advertising” with the possibility of further of confirming or refining them with the analysis of other similar comments.

category_id	post_id	comment_id	comment_parent_id	comment_author_id	comment_date	comment_text	tumb_ups	thumb_down
11	91	703	0	7251	2015-05-20 09:08:19	Martina si fida del supermercato perchè cliente abituale. Avrà sicuramente firmato una liberatoria. Io controllo bene il livello di privacy richiesto e se interessata al servizio sottoscrivo la liberatoria, scarico l'app o richiedo la fidelity card.	0	0
11	91	680	0	7137	2015-05-19 23:08:59	A mio avviso si è fidata per il semplice fatto che era interessata a provare il funzionamento della app e una volta scaricata, probabilmente, è rimasta dimenticata nel telefono assieme ad una miriade di altre app simili. La presenza di sconti è sicuramente un incentivo. Io spesso scarico app e, anche se mi curo di vedere quanto invadano la privacy, a volte le installo ugualmente perchè mi serve o non ho tempo o voglia di cercarne altre	0	0

Figure 3. A printout of an extract from the dataset containing two individual posts with related information.

In a second phase, researchers reconciled the codes they used by merging or renaming some codes. By the end of the process, 11 (codes) topics were identified and grouped into 4 themes.

The first theme, named *Issues*, concerned (potential) issues that participants recognized in the proposed services. It included the following 4 topics: *privacy*, *sharing with others*, *sharing with service provider* and *monetizing*. The *privacy* topic concerns the mentioning of privacy issues (for example, “*I’d be worried about privacy when using this service.*”). The topic of *sharing information* appears in two different contexts: one related to sharing (information) with the service provider and one related to sharing with others (specifically, other users of the service). The *monetizing* topic refers to comments that discussed the possibility of being paid as a compensation to releasing personal data.

The second theme, *Interaction*, concerned aspects related to interaction with the services and included 2 topics: *content* and *modality*. The content dimension referred to possible criticalities expressed with respect to the content of the service provided or the modalities by which it is provided. For example, in discussing personalized advertising of products, the comment “*a possible risk might be that a product is suggested [only] for marketing reasons [rather than for real potential interest of the customer]*” refers to the topic content. The *modality* topic refers to similar criticalities but focused on the modalities in which the service is provided, for

example in the comment “[a drawback might be that] it may prompt unrequested phone calls for useless surveys”.

The third theme, *Efficacy*, was about the perception of services by participants. It included 2 topics: *trust* and *reliability*. The *trust* topic refers to comments dealing with aspects related to how service providers can be trusted by customers when managing or storing their personal data. For example, the comment “in this way, retail companies would push us into buying those products with higher profit margins for them and this would damage small producers” expresses a clear statement about trust in the service under discussion. The *reliability* dimension emerged from comments in which participants expressed how they felt about technology discussion. For example in the comment: “I don’t think that such feature would be useful, in shopping, everyone has to behave according to their own needs and possibilities. Sometimes, you need to choose a cheaper product because you can’t afford a better one.”

In a third phase (the so-called “closed coding” phase in Grounded Theory), each comment was systematically annotated using the topics identified: for each comment, an annotator decided whether each given topic was mentioned or not and if mentioned with a positive or negative meaning. For example, the comment “I’d be worried about privacy when using this service.” would be tagged as negative with respect to the topic privacy and as not applicable with respect to the other topics. This activity was meant to use the topics as a coding scheme in order to provide quantitative data about the content of the discussions. Furthermore, a fourth theme, *Attitude*, was also added to quantify positive or negative attitudes toward the perceived usefulness and the perceived attractiveness of the specific technologies envisioned in the scenarios. We introduced the topics of *attitude toward use* for which, unlike the other dimensions, we considered 4 values: *enthusiast, balanced, concerned and critical (2 positive and 2 negative)*.

For this phase, a manual for annotators was prepared, explaining the rationale underpinning annotation topics, as well as providing guidelines on how to recognize topics in comments and how to tag them. Two annotators were hired to code all the comments. Each annotator coded half of the corpus but a portion (around 20% of the comment) was annotated by both annotators in order to test the inter-annotator agreement (which measures consistency and consequently the soundness of the annotation). Agreement was computed through the Cohen’s kappa statistics and it was sufficiently high on all the dimensions as showed in Table 3.

Table 3. Cohen's Kappa values for inter-annotator agreement on the dimensions of the coding scheme.

Theme	Dimension	Cohen’s Kappa
Issues	Privacy	0.778
	Sharing with others	0.678
	Sharing with provider	0.696
	Monetizing	0.896
Interactions	Content	0.695
	Modality	0.620
Efficacy	Trust	0.588
	Reliability	0.577
Attitude	perceived usefulness	0.811
	perceived attractiveness	0.751
	attitude toward use	0.740

Results

In this section, we present an analysis of the participation and discussion dynamics during the three-week period based on the Owela platform logs and on the content analysis of the posts. A

discussion about the implication of these results as well as some lessons learned for the research community are presented in the next section.

Participation

Overall, 2,822 comments were posted in the period and 3,131 “thumbs” (like or dislike on a comment of another participant) were given.

As explained above, our participants received points for posting comments and for attributing “thumbs” to comments of others. In order to receive the compensation for their participation, they were required to reach at least 30 points in the three-week of the study. Only 6 participants (less than 5% of participants) failed to reach this minimum level of participation. Another 5% (n=7) accrued less than 49 points (which could be considered just above the minimum required to receive the compensation). Still, more than half of participants (53%, 75 people) did contribute up to 4 times the minimum level of participation required and 52 people (37%) provided up to 20 times more activity than required (see Figure 4).

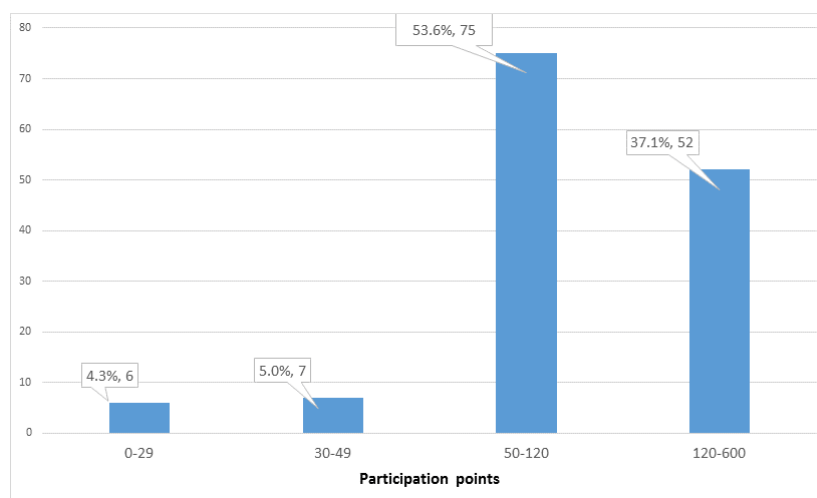


Figure 4. Distribution of participants according to participation points (30 points were the minimum required to receive the compensation for the entire 3-week period).

The majority of participants (n=101, over 72%) commented to all 11 scenarios and over 90% of participants (n=128) commented on at least 7 scenarios.

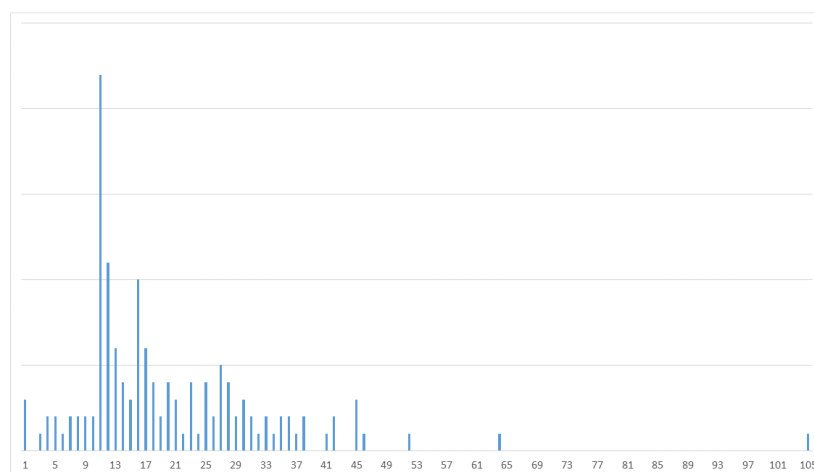


Figure 5. Number of comments posted (horizontal axis) by participants.

The average number of comments per participant was 20 (sd=13.36). Yet, the distribution of comments per participant is quite unbalanced (see Figure 5): 15 participants contributed with less than 10 comments, 74, nearly half of participants, contributed between 10 and 20 comments, and 52 participants contributed with more than 20 comments. One single participant, the top performer, contributed 105 comments.

The average length of the comments was 26.84 words (sd=23.10). Again, the distribution is quite unbalanced: 5% of the comments have less than 20 words and 80% have less than 40 words. The longest comment has 258 words.

Participation Dynamics

In general, participants commented more on the first scenario available and less on the others (as discussed above, the 4 scenarios offered for the week were posted on Monday morning but they were displayed in a certain order on the interface). Figure 6 shows the distribution of the comments for each scenario; the order from left to right reflects the order in which they were available on the interface. For each week, the first scenario to the left received more discussion than the others. Yet, in the second week, the third scenario received a slightly higher number of comments than the last scenarios and a similar pattern was observed for the third week.

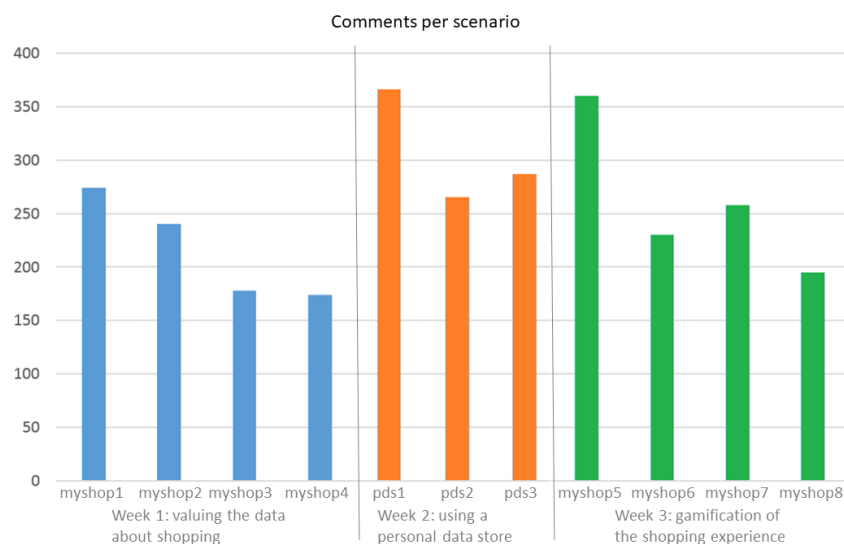


Figure 6. Scenario comment distribution (each bar is a scenario and the order from left to right is the order in which they were available in the interface).

A progressive increase in participation over the three-week period was observed: 866 comments were posted during the first week, 913 (over 5% more) over the second week and 1044 on the third week (over 14% more than the previous week and 20% more than the first week).

The response reaction patterns for participation were similar for the 3 weeks. For each set of scenarios, most of the discussion happened on the week it was introduced, with a clear peak on the first few days. Still, participation distribution in the last week tends to be more balanced than the previous two weeks. As shown in Figure 7, in the first week, 48% of comments were posted on the first day (labelled as 0 in the figure) and 31% on the second day; overall 97% of the comments were received during the week before the proposal of the second set of scenarios. The second week had an even greater peak on the first day, when over 51% of the comments were posted, while on the second day there were overall less comments (68%) than the first week. During the week, we recorded a similar proportion of comments (96%). Therefore, during the second week, contributions tended to be slightly more balanced. The third week, contributions were even more balanced over the week although still more than half of the comments were posted during the first two days.

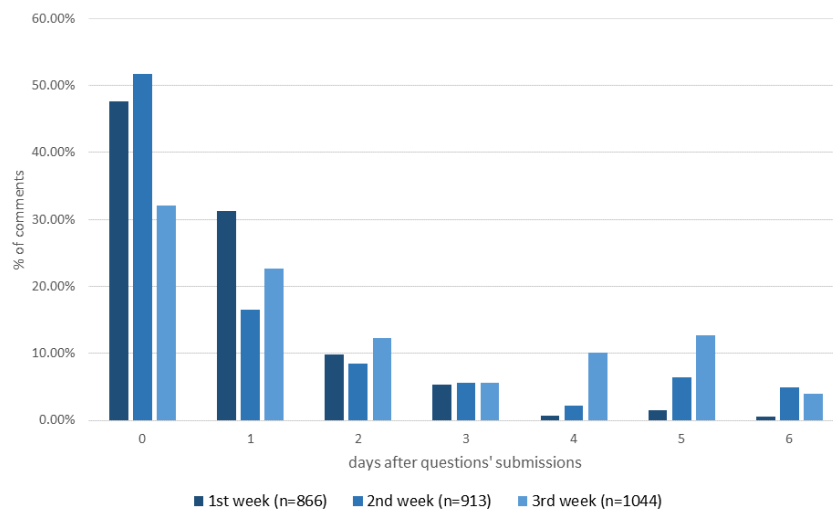


Figure 7. Comment distribution with respect to reaction time after scenario postings (the percentages on the vertical axis refer to the number of comments posted for each week).

Community interaction

To determine whether the participants actually interacted with others or whether they simply contributed their individual comments, we used two measures: we considered, for each participant, (i) the number of comments containing direct replies to questions and (ii) the number of comments to the same question.

Regarding the former, we conducted a quick qualitative investigation on a comment sample. When participants used the “reply” option, they usually meant to comment on others’ comments. Yet, in some cases, comments whose content was clearly a response to previous comments were posted as new posts (that is, without using the reply mechanism). In further analysis, we use the comments posted with the reply mechanism as true replies and assumed that the few “false” replies may be compensated by the reply comments posted as regular comments.

As far as the latter measure is concerned, we assumed that more than one contribution by the same person to the same question was likely to mean an ongoing conversation. In total, 516 comments were entered by using the reply mechanism (about 18% of the 2,823 comments). The percentage of comments, which were replies to other comments changed over the three weeks, similarly to what was observed for comments: 14% (121) of the 866 comments of the first week were replies, 22% (168) of the 913 questions of the second week, and 19% (192) of the 1044 questions of the third week. If we assume that the 520 comments that were replies were equally distributed over the 3 weeks, there is a statistical difference between the actual distribution and the expected one (chi-squared test $\chi^2 = 0.000774352$, $df=2$, $p < 0.01$). Respectively, the first week we observed 24% less replies than expected (equally distributed), the second week 19% more than expected, and the third week 3% more than expected.

An average of 1.11 ($sd=0.18$) commented on the same question by the same participant with a maximum of 2.5 comments by a single person. With respect to response time, the average number of comments per scenario was slightly lower in the first week (mean=1.09, $std=0.14$), it had a slight increment in the second week (mean=1.13, $sd=0.35$) and it went slightly down in the third week (mean=1.10, $sd=0.42$) but the differences are not statistically significant.

Evaluation vs. Ideation Questions

As for the difference in comments posted in response to evaluation questions (namely, those questions that solicit service assessment) with respect to comments posted following ideation questions (those soliciting new ways of implementing a service), no statistical evidence was observed. On average, evaluation questions elicited 67.83 comments per scenario, while creative

questions elicited 65.58 comments (t test not significant). Yet, the creative questions have a wider standard deviation with respect to the evaluation questions (22.76 and 32.28 respectively).

All participants responded to at least one “evaluation” question, but 7 participants consistently skipped all the ideation questions. Indeed, among those 7 participants (4 female and 3 male) were the 5 participants that did not reach the minimum score required to receive the reward.

As far as the length of the comments is concerned, for both types of questions, the difference is statistically significant but not very large: evaluation questions have comments of 27.85 words on average while creative questions have comments of 24.29 words on average ($t=4.0348$, $df=1760.8$, $p\text{-value}=5.699e-05$). In this case, the standard deviation is wider for the evaluation questions ($sd=24.29$) than for the creative questions ($sd=19.56$).

As for replies, the difference in the number of replies between the two types of questions is not statistically significant. The standard deviation is slightly larger in creative questions than in the evaluation questions (for evaluation questions the average proportion of direct replies are 0.18 with a standard deviation of 0.38 while for the creative questions the average proportion is 0.20 with a 0.40 standard deviation).

Participant Contribution

A detailed presentation about the analysis of the actual content is beyond the scope of this paper. Indeed, the ultimate goal of the coding scheme presented above was to provide the companies an initial assessment about design concepts. Just as an example, Figure 8 shows a page of the final report delivered to the companies. It summarized the discussions on the different scenarios both quantitatively (by counting the recurrences of the topic tags) and qualitatively (by reporting the most informative quotes by participants).

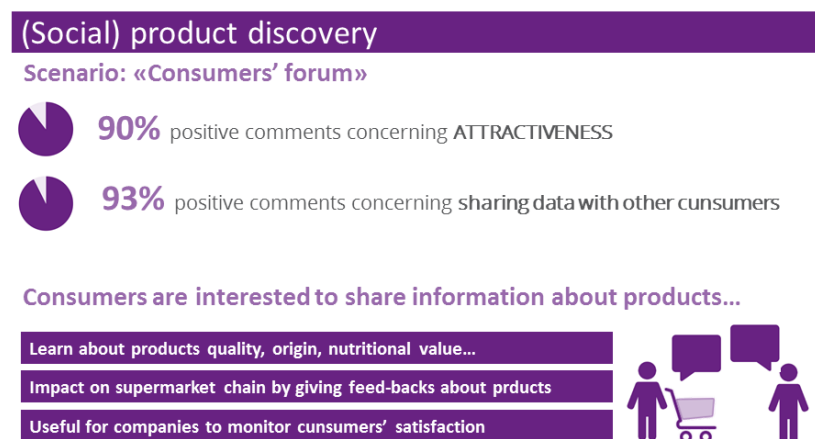


Figure 8. An extract of the comment analysis report delivered to the companies that commissioned the study

Nevertheless, for what concerns the focus of the paper, namely analysis of the participation, the annotated data may provide a better understanding of the quality of the comments gathered during the online discussion.

Table 4 reports the topics annotated for each dimension as either positive or negative. For example, there were 43 comments in which privacy was mentioned both in positive terms (28 cases) and in negative terms (15 cases), while the attitude toward use was by large more intensely discussed (more than 85% of the comments revealed an attitude toward use of a service). This suggests that even if privacy was not discussed much (or at least it was discussed less than other topics), the attitude toward privacy among participants was slightly negative since there were more negative comments than positive ones ($\chi^2=3.93$, $df=1$, $p<0.05$). On the other hand, the attitude toward use is in general positive (enthusiast and balanced scores vs critical and wary scores; $\chi^2=158.37$, $df=3$, $p<0.01$). In any case, it is worth noting that one third of the comments are on the negative side and they contributed to achieve a better understanding of the different points of view on the proposed services.

Table 4. The number of comments annotated for the different themes and topics of the coding schemes (the last column refer to the proportion of comments coded as either negative or positive with the given topic).

		Negative	Positive	Total	%		
Issues	Privacy	28	15	43	1.5%		
	sharingw users	296	329	625	22.1%		
	sharingw provider	110	534	644	22.8%		
	Monetizing	14	174	188	6.7%		
Interactions	Content	10	53	63	2.2%		
	Modality	116	28	144	5.1%		
Efficacy	Trust	84	87	171	6.0%		
	Reliability	60	9	69	2.4%		
Attitude	Usefulness	228	498	726	25.7%		
	Attractiveness	236	783	1019	36.0%		
		Critical	Wary	Balanced	Enthusiast	total	%
	Attitude toward use	603	356	769	687	2415	85.4%

The different numbers of annotations for the topics are partially due to the fact that some scenarios (and some questions) were focused more on some dimensions rather than others. Yet, it might be also due to the fact that some topics were considered more worthy of discussion than others by participants. Table 5 reports the annotation distribution for the topics in the 11 scenarios. For example, the topic sharing with the provider was much discussed during the third week scenarios since all the scenarios about the Personal Data Store design idea brought, of course, this topic of discussion into focus. Yet, it might be interesting to point out how myshop4, a scenario about personalized offers from the store, elicited more discussion about the modality of the service rather than issues about sharing with provider and also just a couple of comments about privacy issues (indeed, the comments about privacy and sharing were all negative).

Table 5. Distribution of annotation of topics for the 11 scenarios.

	myshop1	myshop2	myshop3	myshop4	myshop5	myshop6	myshop7	myshop8	pds1	pds2	pds3	
PRIVACY	21	3	0	2	0	0	0	1	15	1	0	43
SHARING_W_USERS	0	210	162	0	37	1	162	0	48	5	0	625
SHARING_W_PROVIDER	86	2	0	2	0	0	0	0	175	103	276	644
MONETIZING	0	1	0	0	0	0	0	0	14	0	173	188
CONTENT	6	3	0	3	0	0	0	0	0	51	0	63
MODALITY	8	10	1	62	27	0	0	3	15	18	0	144
TRUST	18	6	0	0	0	0	0	67	79	1	0	171
RELIABILITY	6	13	4	1	3	1	1	12	14	14	0	69
USEFULNESS	88	103	45	72	42	70	54	50	75	127	0	726
ATTRACTIVENESS	7	27	113	34	172	126	141	90	130	179	0	1019
	273	239	177	173	359	229	257	194	365	264	286	

Figure 9 shows the distribution of the attitude toward use for the different scenarios. For the sake of clarity, the 4 values have been mapped on a single index with the formula $2 * Enthusiast + Balanced - Wary - 2 * Critical$. In this way, the *Critical* and *Wary* scores reduce the numerical value of the index while the *Enthusiast* and *Balanced* scores push it up; furthermore, the *Enthusiast* and *Critical* scores are computed as twice the *Balanced* and *Wary* scores. Using this index, it becomes apparent how some scenarios were overall judged positively (in particular myshop6 and the pds2) while other rather negatively (myshop2 and myshop5).

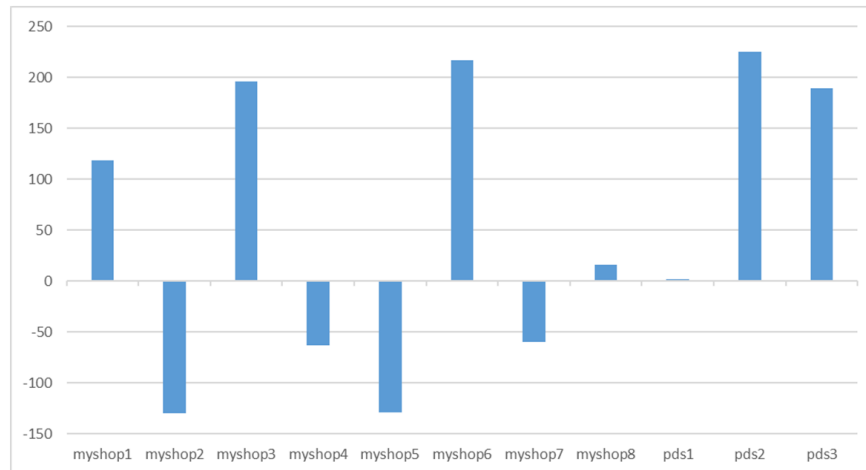


Figure 9. Distribution of a cumulative index for the attitudes toward use ($-2 * \text{Critical} - \text{Wary} + \text{Balanced} + \text{Enthusiast}$).

Table 6 reports the actual counting of the attitudes toward use for each scenario.

Table 6. The distribution of the attitude toward use.

	Critical	Wary	Balanced	Enthusiast	Index
myshop1	19	27	65	59	118
myshop2	91	49	45	28	-130
myshop3	13	7	57	86	196
myshop4	70	19	50	23	-63
myshop5	119	68	97	40	-129
myshop6	2	5	56	85	217
myshop7	90	28	40	54	-60
myshop8	52	28	68	40	16
pds1	83	57	91	67	2
pds2	34	30	71	126	225
pds3	30	38	129	79	189

Discussion and Lessons Learned

Our approach was based on two mechanisms that we believe may foster participant involvement: (i) the use of narrative scenarios and of structuring questions to elicit discussion, and (ii) a reward mechanism that was intentionally constructed on effective, progressive (the compensation was assigned on a weekly basis) and transparent (the participants could check their status in every moment) contribution.

Our results seem to suggest that overall the exercise was successful since we had a high number of comments and expressions of interest (thumbs up or down) in a relatively quick and inexpensive way. This data allowed us to compile a report rich in insights about the concepts of the services proposed by our stakeholders.

In this section, we discuss the research questions presented in the introduction by using the data produced above as evidence.

RQ1: Does this approach foster rich discussion or mainly unrelated individual

contributions?

The comment distribution per participant was highly unbalanced, as usually observed in online discussions (see among others Nielsen 2006; Lampe et al. 2010). Still, in our case, we have evidence that the majority of participants provided regular contributions throughout the study instead of concentrating their attention on few scenarios. Furthermore, an interesting aspect is that most participants contributed much more than the minimum required for obtaining the reward.

Like other studies (for example, Schneider et al. 2012), the comments posted by our participants were overall not very long: the longest one is around half a page but most of them were shorter than a paragraph. Yet, one out of five was a direct reply to another participant's comment and in several cases participants contributed more than once for each discussion. This may suggest a relatively vivid conversation rather than a list of individual short contributions. This might be different from what observed in related works (see among other, Cowley and Radford-Davenport 2011; Schneider et al. 2012; Følstad 2016). We assume that this effect was partially due to the use of scenarios (shorts stories involving fictional characters using the proposed services) rather than descriptions of services since scenarios are more likely to elicit discussions (Carrol and Rosson 1992; Cooper 1999; Carroll 2000; Jenkins et al. 2010).

RQ2: Do the participation and the contribution patterns of the community change over time?

The contribution was overall relatively constant over time and most participants contributed to almost all the scenarios. Contributions gradually increased as the weeks passed. This fact, though, is only partially explained by the slight rise in the number of participants in the three week period. It seems that our participants needed a “warm up” phase in order to get started.

We decided to post all weekly scenarios and probe questions each Monday morning and to leave them online until the end of the study. This turned out to be not so productive. Each week, we had a clear peak in participation the first day with a drop in the following days. The drop was sharper during the first week while it was less pronounced but still quite noticeable during the other two weeks.

This pattern had an effect also on the amount of discussion in the different scenarios. Although the scenarios were not numbered and they were presented at the same time, there was an order in the presentation and this order had a very strong effect: each week, the first scenario most attracted the attention of participants. Yet, the pattern does not seem to be completely order-dependent since for the scenarios after the first set, there are some slight differences in the contribution patterns (see Figure 6). We can assume that even in the presence of a strong “primacy effect”, the specific interest of the scenarios has an effect on the amount of discussions they elicit.

RQ3: Is this approach more suited to assess design concepts or to elicit new ideas?

From our data, we did not find strong evidence that might support that our approach is best suited for either assessing ideas or eliciting new ideas. The number of comments elicited by the different types of questions are similar as are the number of replies they generate. The participants who consistently skipped the creative questions were the less active ones. Indeed, we found that comments to evaluation questions are slightly longer than comments to creativity questions but the small actual difference (3 characters on average) and the lack of other evidence might suggest that we are dealing with a statistical artefact.

Conclusion

In this paper, we presented a case study of an approach to assess early-stage design concepts through an online platform. The study is framed as an explorative case study and it discussed some lessons learned that need to be further investigated. Still, we think that the data collected during the study and the analysis presented in this paper may provide some valuable lessons for replicating this approach.

In particular, we believe that our strategy, based on framing the discussion by means of short narrative scenarios and direct questions used as probes were effective in engaging the discussion as was the use a simple mechanism to link a reward to the actual contribution with a quota-based scheme. Further studies are needed to further investigate these strategies in better controlled settings.

As for our research question on which type of probe questions (evaluative or creative) is more suited for this approach, we did not find evidence that either type of questions is more appropriate since both seem to have elicited the same level of discussion from our participants.

Finally, the analysis of the participation dynamics during the three weeks may provide evidence that the discussion can be kept alive by introducing new topics at regular intervals. We arbitrarily chose to introduce the whole set of scenarios for each given idea once a week and observed a peak of contributions for each set. Yet, the analysis seems to suggest that a shorter interval might be even more effective. Again, further studies are need to better assess this strategy.

References

- Amichai-Hamburger, Yair, Tali Gazit, Judit Bar-Ilan, Oren Perez, Noa Aharoni, Jenny Bronstein, and Talia Sarah Dyne. "Psychological Factors behind the Lack of Participation in Online Discussions." *Computers in Human Behavior* 55 (February 2016): 268–77. Available at <https://doi.org/10.1016/j.chb.2015.09.009>.
- Behrend, Tara S., David J. Sharek, Adam W. Meade, and Eric N. Wiebe. "The Viability of Crowdsourcing for Survey Research." *Behavior Research Methods* 43.3 (2011): 800–813. Available at <https://doi.org/10.3758/s13428-011-0081-0>.
- Bødker, Susanne. "Scenarios in User-Centred Design—setting the Stage for Reflection and Action." *Interacting with Computers* 13.1 (2000): 61–75. Available at [https://doi.org/10.1016/S0953-5438\(00\)00024-2](https://doi.org/10.1016/S0953-5438(00)00024-2).
- Bonner, S.E., R Hastie, G.B. Sprinkle, and S.M. Young. "A Review of the Effects of Financial Incentives on Performance in Laboratory Tasks: Implications for Management Accounting." *Journal of Management Accounting Research* 12 (2000). Available at <https://doi.org/10.2308/jmar.2000.12.1.19>.
- Carroll, John M., and Mary Beth Rosson. "Getting around the Task-Artifact Cycle: How to Make Claims and Design by Scenario." *ACM Transactions on Information Systems* 10.2 (1992): 181–212. Available at <https://doi.org/10.1145/146802.146834>.
- Carroll, John M. 2000. *Making Use: Scenario-Based Design of Human-Computer Interactions*. Boston: MIT Press, 2000.
- Cavaye, Angèle L.M. "Case Study Research: A Multi-Faceted Research Approach for IS." *Information Systems Journal* 6.3 (1996): 227–242. Available at <https://doi.org/10.1111/j.1365-2575.1996.tb00015.x>.
- Cooper Alan. *The Inmates are Running the Asylum*. Indianapolis: Macmillan Publishing Co, 1999.
- Keith K. Cox, James B. Higginbotham and John Burton. "Applications of Focus Group Interviews in Marketing." *Journal of Marketing* 40.1 (1976): 77-80. Available at <https://doi.org/10.2307/1250683>.

- Cowley, Jennifer A., and Julie Radford-Davenport. "Qualitative Data Differences between a Focus Group and Online Forum Hosting a Usability Design Review A Case Study." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 55 (2011): 1356–1360. SAGE Publications. Available at <http://pro.sagepub.com/content/55/1/1356.short>.
- Deci, Edward L. "Effects of Externally Mediated Rewards on Intrinsic Motivation." *Journal of Personality and Social Psychology* 18.1 (1971): 105–115. Available at <http://dx.doi.org/10.1037/h0030644>.
- Friedrich, Pirjo. *Web-based co-design: Social media tools to enhance user-centred design and innovation processes*. PhD diss., Aalto University School of Science, 2013. Available at <https://aalto.aalto.fi/bitstream/handle/123456789/10255/isbn9789513880040.pdf>.
- Følstad, Asbjørn, Ida Maria Haugstveit, Knut Kvale, and Amela Karahasanovic. "Design Feedback From Users Through an Online Social Platform: Benefits and Limitations." *Interacting with Computers* 28.4 (2016): 421–36. Available at <https://doi.org/10.1093/iwc/iwv017>.
- Glaser, Barney G. and Anselm L. Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine de Gruyter, 1967.
- Greenberg, Michael D., Matthew W. Easterday, and Elizabeth M. Gerber. "Critiki: A Scaffolded Approach to Gathering Design Feedback from Paid Crowdworkers." In *C&C '15 Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, ACM Press, 2015. Available at <https://doi.org/10.1145/2757226.2757249>.
- Hagen, Penny. *The Changing Nature of Participation in Design: A practice-based study of social technologies in early design research*, PhD diss., Faculty of Engineering and Information Technology, University of Technology, Sydney, 2011. Available at <https://opus.lib.uts.edu.au/handle/10453/28080>.
- Heyman, James, and Dan Ariely. "Effort for Payment a Tale of Two Markets." *Psychological Science* 15.11 (2004): 787–793. Available at <https://doi.org/10.1111/j.0956-7976.2004.00757.x>.
- Janzik, Lars, and Christina Raasch. "Online Communities in Mature Markets: Why Join, Why Innovate, Why Share?" *International Journal of Innovation Management* 15.4 (2011): 797–836. Available at <https://doi.org/10.1142/S1363919611003568>.
- Jenkins, Nicholas, Michael Bloor, Jan Fischer, Lee Berney, and Joanne Neale. "Putting It in Context: The Use of Vignettes in Qualitative Interviewing." *Qualitative Research* 10.2 (2010): 175–198. Available at <https://doi.org/10.1177/1468794109356737>.
- Iuso, Bill. "Concept Testing: An Appropriate Approach." *Journal of Marketing Research* 12.2 (1975): 228. Available at <https://doi.org/10.2307/3150449>.
- Lampe, Cliff, Rick Wash, Alcides Velasquez, and Elif Ozkaya. "Motivations to Participate in Online Communities." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM Press, 2010. Available at <https://doi.org/10.1145/1753326.1753616>.
- Leonardi, Chiara, Nicola Doppio, Bruno Lepri, Massimo Zancanaro, Michele Caraviello, and Fabio Pianesi. "Exploring Long-Term Participation within a Living Lab: Satisfaction, Motivations and Expectations." In *Proceedings of NordiCHI 2014*, 927–30. ACM Press, 2014. Available at <https://doi.org/10.1145/2639189.2670242>.

- Ma, Xiaojuan, Li Yu, Jodi L. Forlizzi, and Steven P. Dow. "Exiting the Design Studio: Leveraging Online Participants for Early-Stage Design Feedback." In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*, 676–85. ACM Press, 2015. Available at <https://doi.org/10.1145/2675133.2675174>.
- Mason, Winter, and Duncan J. Watts. "Financial Incentives and the Performance of Crowds." In *ACM SIGKDD Workshop on Human Computation (HCOMP '09)* 11 (2010), 100–108. Available at <http://dl.acm.org/citation.cfm?id=1809422>.
- Murgado-Armenteros, Eva María, Francisco José Torres-Ruiz, and Manuela Vega-Zamora. "Differences between Online and Face to Face Focus Groups, Viewed through Two Approaches." *Journal of Theoretical and Applied Electronic Commerce Research* 7.2 (2012): 73–86. Available at <http://dx.doi.org/10.4067/S0718-18762012000200008>.
- Näkki, Pirjo, and Kaisa Koskela-Huotari. "User Participation in Software Design via Social Media: Experiences from a Case Study with Consumers." *AIS Transactions on Human-Computer Interaction* 4.2 (2014): 129–52. Available at <https://doi.org/10.17705/1thci.00043>.
- Nielsen, Jacob. 2006. *The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities*. Available at <https://www.nngroup.com/articles/participation-inequality> (retrieved May 28, 2018).
- Norman Donald and Stephen W. Draper (Eds.). *User Centered System Design*. Hillsdale: L. Erlbaum Associates Inc, 1986.
- Nunamaker, Jay F., Robert O. Briggs, Daniel D. Mittleman, Douglas R. Vogeland, and Pierre A. Balthazard. "Lessons from a Dozen Years of Group Support Systems Research: A Discussion of Lab and Field Findings." *Journal of Management Information Systems* 13.3 (1996): 163–207. Available at <https://doi.org/10.1080/07421222.1996.11518138>.
- Reyes, Lill Francis Miranda, and Sisse Finken. "Social Media as a Platform for Participatory Design." In *Proceedings of the 12th Participatory Design Conference: Exploratory Papers, Workshop Descriptions, Industry Cases – Volume 2 (PDC '12)*, 89. ACM Press, 2012. Available at <https://doi.org/10.1145/2348144.2348173>.
- Rosson, Mary Beth, and John M. Carroll. "Scenario-Based Design." In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, 1032–50. Lawrence Erlbaum Associates, 2002.
- Ryan, Richard M., and Edward L. Deci. "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions." *Contemporary Educational Psychology* 25 (2000): 54–67. Available at <https://doi.org/10.1006/ceps.1999.1020>.
- Schneider, Sid J., Kerwin Jeffrey, Frechtling Joy, and Vivari Benjamin A. "Characteristics of the discussion in online and face-to-face focus groups". *Social Science Computer Review* 20.1 (2002): 31-42. Available at <https://doi.org/10.1177/089443930202000104>.
- Eric Von Hippel. "Lead Users: A Source of Novel Product Concepts." *Management Science* 32.7 (1986): 791–805. Available at <https://doi.org/10.1287/mnsc.32.7.791>.
- Yen, Yu-Chun (Grace), Steven P. Dow, Elizabeth Gerber, and Brian P. Bailey. "Social Network, Web Forum, or Task Market? Comparing Different Crowd Genres for Design Feedback Exchange." In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*, 773–84. ACM Press, 2016. Available at <https://doi.org/10.1145/2901790.2901820>.

Yin, Robert K. *Case study research: Design and methods (applied social research methods)*. London and Singapore: Sage, 2009.

Yndigejn, Signe L. "Extending Design Encounters with Use of Social Media." In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction (OZCHI '10)*, 356. ACM Press, 2010. Available at <https://doi.org/10.1145/1952222.1952301>.

Zeng, Michael A. "The Contribution of Different Online Communities in Open Innovation Projects." In *Proceedings of The International Symposium on Open Collaboration*, 9. ACM, 2014. Available at <http://dl.acm.org/citation.cfm?id=2641593>.

Appendix

The scenarios and corresponding probes questions used during the study. All the material has been translated from Italian.

SCENARIOS	PROBES QUESTIONS
<p>ID: myShop1 Title: My purchases Martina has just downloaded the new "MyShop" app that allows her to visualize all the purchases she does on her usual retail chain stores. Martina feels she spent more than usual this month. She looks on the app the graph of the monthly expenses and she realized that she has really overspent this month with respect to the previous ones. For each shopping session, the graph also shows the division by product category (groceries, home products, etc.). Martina understands why she overspent the last month: over 20% of the groceries she bought were organic...</p>	<p><u>Q1. Let's talk about usefulness of the app.</u> In the story, Martina sees the trend of her purchases by category and type and she can analyze her expenses. Do you think that this information can change how Martina actually shops? What about your habits, would your habits change using such a system? <u>Q2. Why do Martina trust the store? What would you do?</u> In your opinion, what convinced Martina to trust and agree to share the information about her expenses with the retail chain? Would you trust them? What would you need to trust them? <u>Q3. Can you think of other uses?</u> If you were Martina, which other information would you like to see? Can you think of other functionalities that you would expect on this app? <u>Q4. Which kind of risks is Martina exposed to?</u> Which kind of risks you can see in the sharing of information about purchases with a retail chain?</p>
<p>ID: myShop2 Title: Compare my purchases The app (see story "My purchases") allows Martina to compare her purchases with those of other people similar to her. In order to use this functionality, Martina had to agree to share her data with other customers (in an anonymous way). Thanks to this comparison, Martina realized that she spends more than 5% more than the customers similar to her in organic products and around 7% less in home products.</p>	<p><u>Q1. Let's talk about usefulness of the app</u> How much useful you consider a service that shows you how much consumers similar to you spend? In your opinion, is such a system able to change the way you shop? <u>Q2. What do you expect more from a similar service?</u> Which visualization/graphs would you expect from a similar service? <u>Q3. Which kind of risks is Martina exposed to?</u> Which risk is Martina exposed to n sharing own personal information? <u>Q4. Which kind of data would you share?</u> And how would you feel in sharing such personal data? Which data would you not share? <u>Q5. More advantages or more risks?</u> Do you think that Martina has more advantages or disadvantages in sharing own personal data? <u>Q6. How would such a service change the way Martina does her grocery shopping?</u> Do you think people would change they do their grocery shopping if they compare their shopping data with similar people?</p>

<p>ID: myShop3 Title: Consumers' forum Martina – with her new app – can become part of a consumers' online community. She immediately uses the app to view other consumers reviews about some new products that she still never bought. Once she tried a product, she write a review of that product. She found that product very scarce and she complaint about this.</p>	<p><u>Q1. Is it interesting?</u> How Interesting do you consider a service that allows customers to share information about products? <u>Q2. Which are the advantages and disadvantages?</u> Could you imagine possible advantages or disadvantages for customers? Which ones? And for the supermarket chains, what can be the advantages or disadvantages?</p>
<p>ID: myShop 4 Title: Personalized offers The app also recommends to consumers specific products based on the purchases made in the past and some socio-demographic information. To Martina the app suggests products such as organic bread, goat cheese, carrots bio. To Paul, a student at the university, the app recommends different products: Knorr soups, tomato sauce, canned beers</p>	<p><u>Q1. Let's talk about the service!</u> What do you think about a similar service? Which are the advantages? <u>Q2. Which useful services and information could a similar app can provide for you?</u> <u>Q3. Which are the risks consumers are exposed to?</u> In the scenario above, Paul and Mary receive very different recommendations. Which are the risks to be profiled and receive personalized offers?</p>
<p>ID: myShop 5 Title: Treasure hunt Michael just downloaded the app MyShop. Back to home, the app suggest to Michele that there is a game to play: "A treasure hunt of a mysterious product!". The game consists in finding the product in the store following a number of clues. Michele understands that if he found the product, he has to take a picture of it and send the image in order to gain points and play against other consumers.</p>	<p><u>Q1. How would you define this service?</u> What three adjectives would you use to describe this app? <u>Q2. Let's talk about the limits!</u> Which problems do you see related to playing this game during your shopping? <u>Q3. Game duration.</u> In your opinion, how long should the treasure hunt last? In how many games would you like to participate in a year? How often would you like to be invited to a new treasure hunt game? <u>Q4. Motivations.</u> In this scenario, Michael accepts to participate to the game during his shopping activities. What would motivate you to participate? Why wouldn't you participate? <u>Q5. Creativity space.</u> How could an intelligent app that knows your spending habits could propose/give you advice in a fun way new products to try? <u>Q6. Could the app be more "social"?</u> In what ways would you</p>
<p>ID: myShop 6 Title: Let's classify the products Michael just finished his grocery shopping. Back home, he receives a notification. The supermarket chain proposes to Michael to be engaged in a new campaign that addresses environmental issues. Michael is asked to categorize products according to products origins, packaging, weight, etc. This data collection will help the supermarket chain to collect useful information I order to do more sustainable choices and to promote more "green" behaviors. In return for his help, the supermarket gives Michael points to be used for his shopping.</p>	<p><u>Q1. How would you define this service?</u> What three adjectives would you use to describe this app? <u>Q2. Let's talk about the service!</u> In this scenario, Michael participated in the classification of the products because he can earn points to be used as discounts or other prizes. What do you think? <u>Q3. Let's talk about motivations and incentives!</u> Which types of incentives could push you to dedicate some time to classify products? For which reasons would you participate to such a campaign?</p>
<p>ID: MyShop 7 Title: Consumers online community As in the previous user story, Michael just finished his grocery shopping. Back home, he receives a notification. The supermarket chain proposes to Michael to be engaged in a new campaign that addresses environmental issues. Differently from the previous scenario, each participant's contribution will be visualized within the app. In the last years, Michael has become more and more aware about sustainability challenges and he is happy to give his support to these kind of initiatives. He start inserting product data and he also decides to add other information, such as the recipe he just did with the products bought! In the profile page, he now notices that he is one of the top contributors. Other participants thanks Michael for sharing the recipe.</p>	<p><u>Q1. How would you define this service?</u> What three adjectives would you use to describe this app? <u>Q2. Let's talk about the service!</u> In this scenario, Michael participates to the classification of the products NOT for receiving points but because he believes it is important to do it and he obtains a social recognition among the community. What do you think? <u>Q3. Other ways to use the app.</u> In these two scenarios, Michael uses the app at home. Are there any other contexts in which Michael classify the products? For ex. waiting in the queue. What do you think? Do you have other options? <u>Q4. Comparing the two MyShop scenarios.</u> Think about the two scenarios we have presented on products' classifications: what could motivate you to participate in products' classification activities?</p>

<p>ID: myShop 8 Title: <i>Tell me what you eat and I will tell you who you are</i> Michael is queuing up at the counter. A notification arrives. A game is proposed to Michael. The game consists in composing a meal following own preferences. The game then gives scores according to how much the meal is healthy. Michael plays and compose a meal with sausage and tomato. The app gives a ow score to Michael but also suggests how to improve the quality of the meal. For instance Michael can replace the sausage with zucchini. Finally, the app gives Michael hints for an healthy diet.</p>	<p><u>Q1.How would you define this service?</u> What three adjectives would you use to describe this app? <u>Q2.Other ways to use the app</u> In these two scenarios, Michael uses the app to get suggestions about healthier choices. Do you have other ideas related to this topic?</p>
<p>ID: PDS 1 Title: <i>Personal data management</i> Mark ha subscribed to the Personal Data Store service, through which he can automatically collect his personal data in a private account. Data such as position (through his smartphone GPS), his social interaction history (calls, messages etc.), physical activity data (through a wristband), data related to his expenses (from his credit cards and bank account) are all available in the Personal Data Store account. Nobody can use these data without Mark's permission. Moreover, the Personal Data Store allows Mark to access an app marketplace where he can activate apps that use his personal data in a controlled and transparent way. Those apps can help Mirko to be more aware about his behavior, increase his well-being and help him save money and so on. Mark can control which apps to activate, which data those apps have access to and how frequent and deactivate apps at any time.</p>	<p><u>Q1.Let's talk about the service:</u> How do you feel about a service like the PDS that can be used to activate trusted services using your personal data? What is the most interesting aspect of the service? What would you be ready to pay for the service? <u>Q2.And the apps based on the PDS?</u> How should the applications function when they use your data? What information would you like on their activity and use of your data? <u>Q3."Social Comparison" app:</u> With the PDS, you can compare your actions with people similar to you by sharing parts of your data anonymously. How do you feel about this kind of opportunity? How do you believe using the application would change the way you do things? <u>Q4.Personalized suggestions:</u> What kinds of application suggestions, based on your own data, would you want the PDS to make? <u>Q5.Let's talk about trust!</u> What (public or private) entity do you feel would be a trustworthy PDS service provider?</p>
<p>ID: PDS 2 Title: <i>The offer of your dreams!</i> Mary has subscribed to the Personal Data Store and she is interested to buy a new car. Through the PDS she can access an app which allows her to receive anonymously personalized offers from some car sellers in her city. Mary decides which data to share with the car sellers: some are automatically extracted (such as statistics on her movements by car) and some are provided by herself. Based on this data the app fills in a request to be sent to the car sellers affiliated to the PDS. The car sellers elaborate Mary's request and send her an offer in a completely anonymous mode. Mary can now decide whether to contact a car seller and remain anonymous for the others who sent her offers.</p>	<p><u>Q1.Advantages:</u> What advantages would an application like this have? How interested would you be in using the application when you make important purchases? <u>Q2.On which types of products?</u> What kinds of purchases would the application suit best? <u>Q3.How much would you pay?</u> What would you be ready to pay for the service?</p>
<p>ID: PDS 3 Title: <i>Make money with your personal data!</i> Michael subscriber of the PDS is interested to have also some economic benefits by sharing his personal data in an anonymous and aggregated mode with public or private organizations. The PDS allows Michael to activate an app through which he can choose which data to monetize. The app is an intermediary for organizations interested in having citizens' data for statistics and profiling. Michael will receive an economic benefit in return, based on the quantity of data effectively shared with those organizations. He then decides to activate the app and choose the types of data to be monetized, being also aware that he can decide to stop sharing his data at any time.</p>	<p><u>Q1.Let's talk about payment!</u> What would you like to receive as a compensation for selling your data? How would you like to receive your remuneration? <u>Q2.Sharing data with whom?</u> How would you like to pick the organizations with which you share your data? How should the organizations be classified in the application? <u>Q3.Free data.</u> On what terms, to which parties and purposes would you be ready to share your personal data free of charge? What data would you be ready to share free of charge for the common good? <u>Q4.Research.</u> What kinds of research projects and surveys would you be interested in participating through the application, if you would be rewarded for your participation?</p>