

アンケート調査の参加モチベーションを向上する質問順番

Improving Motivation in Survey Participation by Question Reordering

シン ロヒット クマール¹ スッパキットパイサーン ウォラポン^{1*}
Rohit Kumar Singh¹ Vorapong Suppakitpaisarn¹

¹ 東京大学院情報理工学研究科コンピュータ科学専攻

¹ Department of Computer Science, Graduate School of Information Science and Technology,
The University of Tokyo

Abstract: We organized an experiment to show that survey participants take part more when the questionnaires started with less aggressive questions. In our earlier work, we used Bayesian probability and graph algorithms to find relative values of each personal attribute. Using that valuation, we created two sets of the questionnaire each differs in question order and ask 33 personal attributes from participants. The first set of the questionnaire ordered from personal attributes with high valuations such as passport number, driving license number, last name, and monthly income to personal attributes with low valuations such as nationality, gender and office country. On the other hand, the second set of questionnaire ordered from those with low valuations to those with higher valuations. As a result, the number of participants who received the second set of the questionnaire and agrees to submit some information is 71.42% more than those who received the first set of the questionnaire. Moreover, the second set of participants spend much less time in filling the questionnaire but provides 1.78% more information on average.

1 Introduction

Personal information in this modern data driven age is critical asset for all business. Often businesses collect users personal information through online registration forms. For example, e-commerce sites, mailing services, social networks, microblogging sites, location, news and weather services, handheld devices etc. All of these required users to register before accessing their services. Businesses need personal information not only to validate genuine users but also for their future planning, product or service feedback, targeted advertisement and marketing as well as personal information trading with third parties. So, directly or indirectly all businesses need users personal information to be competitive and cater their users more effectively and efficiently.

Despite the above fact, collecting personal information is more stringent because, day by day people are

more aware about web and personal information security. To lure the users, companies often provide monetary gifts in exchange of their personal information [4].

In this study, we will focus on a more practical way of collecting user personal attributes, .i.e., an online questionnaire. Online questionnaire surveys are quick and cost effective way of collecting data from large number of people. But often, people feel insecure in providing their personal attributes and leave such questions unanswered. It significantly affects the overall response to the survey. For example, the unanswered question are personal phone number, passport number, driving license number, email address, mobile phone number, date of birth etc. Without these personal information attributes, companies are deprived of many business opportunities whereas, users are forfeit from better services and offers.

From the above discussion, online questionnaire survey looks more promising, if we encourage partici-

*東京大学院情報理工学研究科コンピュータ科学専攻
〒113-0033 東京都文京区本郷 7- 3- 1
E-mail: vorapong@is.s.u-tokyo.ac.jp

participants to reveal their personal information attributes more. We organized an experiment to show that, participants get involved more when survey questionnaire starts with less aggressive questions first. From our previous work, which was based on Bayesian probability and graph algorithms, we defined the relative values for each personal attribute [7]. Utilizing that result, we constructed two sets of survey questionnaire, each differs in question order and urged 33 personal attribute from survey participants. Set-1, start with high valuation most aggressive question of passport number, driving license number, last name and ends with personal attributes which are least aggressive and having low valuation such as nationality, gender and office country. On the other hand, set-2 of survey questionnaire ordered from questions with low valuations to those with higher valuations.

In our experiment, along with user data, we also recorded time of entry to study participants response behavior for each set. From user data we measured average participation and average number of attributes filled for each set. Furthermore, using time we calculated average time taken to submit each set of survey as well as number of times users distracted. Our result indicates that there is an 71.42% increase in submission for set-2 in comparison to set-1 as well as average time taken by participants for set-2 submission is 61.18% less than set-1.

2 Related Works

Here we argued existing research work in the field of questionnaire survey. In section 2.1, effects of question order in a survey is covered and in section 2.2, personal attribute valuation study for an online questionnaire is explained.

2.1 Effect of Question Order

Survey questionnaire order was studied since 1939 [8]. In [5], they performed a study to understand the question order effect on questionnaire survey response. Their questionnaire response was evaluated based on an overall general question followed by series of specific questions related to a particular topic and vice-versa. For example, the general question on current education system asked “How would you describe

the current scenario in higher education system: as extremely good, just OK, or required improvement?” Users were also asked about related specific questions like, research grant shortage in higher education, plagiarism and increasing women participation.

The author also argued that participant response varies more by evaluation capability and users interest about subject of questionnaire than questionnaire order itself. Sometimes asking more specific question first, provides concrete idea about the subject and allow participant to create a relationship in responding to subsequently general questions. This is true when user requires to express his interest. For example, “How often one plays outside and whether or not one has attended any sports event in the past seven days” are objective indications of one’s sports interest. Whereas on the other hand above is not true when response depends upon evaluation capabilities of the participant. For example, the specific question “How about the vegetarian food: Do you think more people will get inclined towards it over the next ten years, or will it decline?” increases the saliency of health consciousness and it requires user to evaluate its response. Asking this specific question first will not strengthen the relationship to later general questions.

Hence interest seems prone to questionnaire order but evaluation does not [5]. Above study interviewed peoples over telephone for their survey, but an online questionnaire survey is different in terms of providing more freedom to participants in submitting their response.

2.2 Valuation of Personal Attributes

Researchers developed a technique of evaluating value of each personal attribute [6]. Firstly they segregated the collected data in two categories, “protected” and “disclosed”. Using Bayes’ formula, they calculated posterior probabilities of protecting or disclosing one attribute when other is already protected or disclosed.

$$P(x|y) = \frac{P(y \cap x)}{P(y)}$$

Where $P(x|y)$ is the probability of protecting attribute x when user already protected attribute y . $P(y \cap x)$ is the probability of protecting attribute x

and y . $P(y)$ is the probability of protecting attribute y . Next, they constructed a directed graph by removing the edges having probability $P(x|y) \leq 0.95$, and considered only on pair of edges having high probability which implies strong relationship between two personal attributes. Each node represent a personal attribute and each edge directed from node X to node Y is regarded as attribute x is more valuable than attribute y . In the next step, they suggested to use graph mining technique to obtain communities to understand disclosure relationship among attributes. This method is useful in understanding the relationship between attributes. In spite of that, constructed graph becomes cumbersome when attribute list increases to many folds. It also hard to construct a hierarchy of top valued to least valued attribute out of it.

In a recent study, researchers carried forward the previous work and proposed the solution of making this large graph useful [7]. They converted the large cumbersome graph into a tree structure. The root node of the tree is the attribute having highest out-degree and least in-degree whereas the leaves of a tree consist of node with highest in-degrees and lowest out-degrees. Each edge in a tree contains the calculated probability of protecting parent node when the child node is protected, i.e., $P(x|y)$. From the constructed tree, they proposed a method of valuation of personal attribute disclosure (VD).

$$VD_X = \prod_{I=1}^{X-1} W_{I,I+1}$$

Where X is the destination node, I is the root node, $V = (I, I + 1, \dots, X - 1, X)$ where I has a path to X which is $I, I + 1, \dots, X - 1, X$ and $W_{I,J}$ is the probabilistic weight of edge $e_{I,J}$. Here J is the subsequent intermediate node $I + 1$. This proposed tree structure and valuation method depicts the hierarchy of personal attributes. Higher VD signifies that participant will protect that personal attribute.

3 Research Methods

We have used previous study for creating our survey questionnaire list [7]. In section 3.1 we described methods used for data collection and in section 3.2 we have provided our questionnaire list.

3.1 Data Collection

For data collection we used an online questionnaire which is developed using node.js¹, express.js² and HTML³. There are two instances of the web server running in parallel and hosting two different questionnaires. First set of questionnaire starts with high valuation personal attributes, i.e., “Passport Number” and ends with least aggressive question of “Office Country”. On the other hand second set of questionnaire starts with less valued personal attribute of “Office Country” and ends with high valued “Passport Number”. We selected 44 people, most of them are students in The University of Tokyo for our online survey and invited through personal email id’s. Half of the recipients received first set of questionnaire and another half received second set of questionnaire. The choice of survey to the recipient is randomly selected by us and respective survey link is provided in their invitation email.

In order to avoid doubt about the organizers of the survey or let them to believe our mail invites as “Spam mail”, we purposefully invited personally known people. We also assured in the invitation about users privacy and data security. In the invitation email content we avoided providing much information to restrict users from any kind of influence. The email content looks as given:

Dear XXX,
 We are a group of researchers at XXX Lab in Department of Computer Science (University of Tokyo). We are collecting users personal information for our research. We are happy to receive your complete input for our form, but feel free to leave the uncomfortable fields. Survey Link: <http://xxxxxx-xxx.net:1234/> Note: Collected data is only used for research and will not be shared with any third party.

3.2 Questionnaire Format

Our survey questionnaire consists of all 33 questions used for male participants in previous study [7] because most of our survey participants are male. Except two questions, we kept remaining 31 questions

¹<https://nodejs.org/en/>

²<http://expressjs.com/>

³<http://www.w3.org/TR/html/>

intact. We replaced question “ID Number” and “Picture” with “Passport Number” and “Driving License Number” respectively because previous study was organized in Thailand where “ID Number” is provided to every citizen but same is not available in Japan and we used “Driving License Number” instead of “Picture” because we assumed driving license is more accurate information about a person than its picture in an online survey about personal attributes.

We also used the same values for personal attribute disclosure(VD) from the previous study [7]. For the simplicity of our research, we carried same VD for the newly added attributes “Passport Number” and “Driving License Number ” as calculated for existing attributes “ID Number” and “Picture”. Our first set of questionnaire composed of sequence given in Table 1, and second set is exactly the reverse of it so that it has less aggressive question first. We invited people for our online survey on September 16, 2015 and collected data until our abstract submission on September 23, 2015. A sample image of our online survey form is provided in Figure 1.

Hello Everyone!!!

We are a group of researchers at Imai Lab in Department of Computer Science (University of Tokyo). We are collecting users personal information for our research. We are happy to receive your complete input for our form, but feel free to leave the uncomfortable fields.

Take Survey

Passport Number

Driving License Number

Last Name

First Name

Middle Name

図 1: Online Survey Form (Set 1)

Rank	Attribute	VD
1	Passport Number (ID Number)	1.0000
2	Driving License Number (Picture)	0.9908
3	Last Name	0.9896
19	First Name	0.9430
21	Middle Name	0.9392
4	Home Address	0.9826
5	Home City	0.9826
6	Monthly Income	0.9789
7	Home Phone	0.9753
8	Office Email	0.9750
9	Highest Education	0.9692
10	Age	0.9583
11	Office Phone	0.9575
12	Marital Status	0.9540
13	Nickname	0.9504
14	Mobile Phone	0.9498
15	Personal Fax	0.9495
16	Number of Children	0.9490
17	Office Address	0.9444
18	Home Zip Code	0.9442
20	Home Province	0.9413
22	Birth Date	0.9383
23	Office City	0.9380
24	Office Zip Code	0.9377
25	Blood Type	0.9336
26	Personal Website	0.9302
27	Office Province	0.9297
28	Personal Email	0.9286
29	Home Country	0.9181
30	Nationality	0.9157
31	First Language	0.9111
32	Gender	0.8992
33	Office Country	0.8969

表 1: Survey Questionnaire Set

4 Results Analysis and Discussion

To test our hypothesis that asking less aggressive question first will increase the submission percentage in an online personal attribute questionnaire, we performed an online survey with two sets of same questionnaire on 44 participants. Order of questions in set-1 is same as provided in Table 1, and set-2, is exact reverse of set-1. Each set is equally divided into

two sets of 22 participants through email invites. Set-1, was submitted by 7 participants; whereas, set-2, was submitted by 12 participants out of 22 participants in each case. Obtained data is analyzed with χ^2 test and Welch’s t-test. Our analysis is provided in Table 2.

	Average Value	Set-1	Set-2	Significance
1	Participation Ratio	7/22 = 0.3181	12/22 = 0.5454	0.127
2	Average time taken for Form Submission	745.71 seconds	289.45 seconds	0.053
3	Average No. of Attributes Filled ⁴	21.86	22.25	0.453
4	Average No. of Attributes Filled ⁵	6.95	12.13	0.071
5	Average No. of Times Distracted	2.86	2.16	0.326

表 2: Comparison between Set-1 (Higher Valuation Attributes to Lower) and Set-2 (Lower Valuation Attributes to Higher)

In row 1 of Table 2, we analyzed survey participation ratio for each questionnaire set. Our experiment result showed that there is an increase of 71.42% more submissions for set-2. It also suggests that, on-line questionnaire survey on personal attributes which start with the least aggressive question first, encourages participants more for submission than questionnaire starts with the most aggressive question first. However, χ^2 test, yields p-value=0.127.

The average amount of time spent by participants in completing our survey for set-2, is 61.18% less than set-1, as given in row 2 of Table 2. This implies that asking, most aggressive questions first, makes users more insecure for the whole survey and that leads to increase in time spend for survey. On the other hand,

⁴For this calculation, we do not considered participants who does not submitted any information.

⁵For this calculation, we considered participants who does not submitted any information as a person who give us zero information.

Welch’s t-test produced p-value=0.053.

In row 3 of Table 2, we calculated the average number of personal attributes responded by participants for each set of questionnaire. Our result showed that there is an increase of 1.78% for set-2 as compared to set-1. Welch’s t-test for significance results, p-value=0.453. Furthermore, when we considered total invitees for each set, i.e., 22 in our calculation, our result improves to 74.53% for set-2 over set-1 and Welch’s t-test for significance improves p-value=0.071, as showed in row 4 Table 2.

Similarly, in row 5 of Table 2, we calculated the average number of times participants distracted in answering a question for more than 30 seconds. Our result hinted that, participants for set-2 are 24.48% less distracted than set-1. This outcome suggests that questionnaire set-2 makes participants more relaxed and hence takes less time in framing answers. On the other hand, set-1, forced people feel unguarded and become extra cautious in revealing their informations and that leads to more distraction in framing answer. However, Welch’s t-test significance produced p-value=0.326.

All above five results, showed improvements for set-2 in comparison to set-1. In addition, at the significance level of 0.10, calculated p-value of 0.053 for the result of row 2 in Table 2, is significant. Moreover, at this significance level of 0.10, the result obtained in row 4 Table 2, with the consideration of total invitees for each set, i.e., 22, obtained p-value=0.071, is also significant. However, significance test does not support our improvements at a significance level of 0.01. We also believe that, for a justifiable significance test, a large number of participants is crucial. However, obtaining such a critical number of participants is very difficult for these kind of experiments.

	PN	LN	HAdd.	HC	MI
Set-1	1	7	6	7	6
Set-2	3	8	7	11	9

表 3: Top 5 Significant Attributes

We also tried concluding top 5 most significant attributes from both sets of questionnaires which users may intentionally avoided to reveal and having higher valuation of personal attribute disclosure (*VD*) value.

These are, *Passport Number (PN)*, *Last Name (LN)*, *Home Address (HAdd.)*, *Home City (HC)* and *Monthly Income (MI)*. Data obtained for each set for these attributes is provided in Table 3. Out of 7 respondents for set-1 and 12 respondents for set-2, very few have revealed their *Passport Number*. However, *Home City* and *Monthly Income* is answered by most of them. In addition, *Last Name* and *Home Address* is almost equally answered by both group of respondents. Despite difference in their ordering, in our opinion, the reason for not having any big difference in the response for each set of above high value attributes is; for set-1, participants are more careful in revealing their most important information; whereas, in set-2 participants were already exhausted in providing initial least aggressive questions and lost their motivation for disclosing higher *VD* personal attributes.

5 Conclusion

For this study, we adopted a psychological technique called “foot in the door” [1], [2], [3]. In this technique, by answering the least aggressive questions first, participants are obligated to answer later comparatively higher valuation aggressive questions. Our result analysis reveals that set-2 questionnaire submission is approximately seventy percentage higher than set-1. Moreover, the average time spent by participants on submission of set-2 is around sixty percentage less than set-1. These results confirms our previous study of personal attribute valuation. It also support our claim that participants respond more in an online questionnaire when it starts with less aggressive questions first. However, significance test support our claim partially. We believe that, with a large set of participants our results significance could improve. As a result, for future works, we are planning to replicate our experiment on a large population to justify our results significance.

参考文献

- [1] Jerry M Burger. The foot-in-the-door compliance procedure: A multiple-process analysis and review. *Personality and Social Psychology Review*, 3(4):303–325, 1999.
- [2] Jonathan L Freedman and Scott C Fraser. Compliance without pressure: the foot-in-the-door

technique. *Journal of personality and social psychology*, 4(2):195, 1966.

- [3] Nicolas Gueguen, Sebastien Meineri, Angelique Martin, and Isabelle Grandjean. The combined effect of the foot-in-the-door technique and the “but you are free” technique: An evaluation on the selective sorting of household wastes. *Ecopsy-
chology*, 2(4):231–237, 2010.
- [4] Steve Kroft. The data brokers: Selling your personal information. retrieved october 22, 2015, from <http://www.cbsnews.com/news/data-brokers-selling-personal-information-60-minutes/>.
- [5] Sam G McFarland. Effects of question order on survey responses. *Public Opinion Quarterly*, 45(2):208–215, 1981.
- [6] Ake Osothongs, Vorapong Suppakitpaisan, and Noboru Sonehara. Evaluating the importance of personal information attributes using graph mining technique. In *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication*, number 104. ACM, 2015.
- [7] Ake Osothongs, Vorapong Suppakitpaisan, and Noboru Sonehara. A proposed method for personal attributes disclosure valuation: A study on personal attributes disclosure in thailand. In *Proceedings of the 9th International Conference on Information Technology and Electrical Engineering (ICITEE 2015)*, pages 409–413. IEEE Xplore, 2015.
- [8] Jeanette Sayre. A comparison of three indices of attitude toward radio advertising. *Journal of Applied Psychology*, 23(1):23, 1939.