Perceived Frequency of Advertising Practices

Sai Teja Peddinti
Google Inc.
psaiteja@google.com

Allen Collins
Google Inc.
allencollins@google.com

Aaron Sedley
Google Inc.
asedley@google.com

Nina Taft
Google Inc.
ninataft@google.com

Anna Turner
Google Inc.
annaturn@google.com

Allison Woodruff
Google Inc.
woodruff@google.com

ABSTRACT

In this paper, we introduce a new construct for measuring individuals' privacy-related beliefs and understandings, namely their perception of the frequency with which information about individuals is gathered and used by others for advertising purposes. We introduce a preliminary instrument for measuring this perception, called the Ad Practice Frequency Perception Scale. We report data from a survey using this instrument, as well as the results of an initial clustering of participants based on this data. Our results, while preliminary, suggest that this construct may have future potential to characterize and segment individuals, and is worthy of further exploration.

1. INTRODUCTION

Characterizing individuals' privacy-related attitudes and behaviors, and segmenting individuals' into groups based on these characterizations, are long-sought-after goals, e.g. [3]. Nonetheless, effective measures and segmentations have been elusive. Consider for example Westin's Privacy Segmentation Index, which is probably the best-known and most often used characterization and segmentation of users according to their privacy-related attitudes [10]: concerns have long existed regarding the assumptions underlying this index and its apparent lack of predictive power, as summarized for example in [22].

Ongoing research in this area follows many arcs. For example, some work seeks to characterize the relationship (or apparent lack thereof) between different measures, e.g. [17, 4, 22]. Other work seeks to produce new constructs and instruments which might yield additional insight and predictive power, e.g. [9, 20]. The work presented in this paper falls in this latter category.

Specifically, we propose a new construct for measuring one aspect of people's privacy-related beliefs and understandings: their perception of the frequency with which information about individuals is gathered and used by others. In this paper, we focus predominantly although not exclusively on web based activities. Intuitively, this construct may be an explanatory factor for privacy-related attitudes and behaviors. For example, if individuals feel that certain practices occur more frequently, those individuals may have higher levels of concern regarding privacy and/or take more actions to protect their privacy, either in relationship to those specific practices or in general.

As a first effort to explore this construct in a concrete way, we have further narrowed the problem space and designed an instrument to measure people's perceptions of the frequency with which information about individuals is gathered and used in advertising practices. By now, many web users have a general awareness about online advertising practices and tracking technology. However these practices are not the same in all applications, nor for all services, and not for different types of data. As previous work suggests [19], it is quite difficult for an ordinary web user to understand how the advertising ecosystem really works, to know about the various business relationships between the different players, and the implications of different technologies used. Hence, in this study we aim to explore how users perceive this ecosystem. In particular, we seek to understand how frequently participants think various practices do or do not occur, and if they have different perceptions for different web services (such as news, email, shopping, social, photos, etc).

We designed a simple metric of people's perception of the frequency of various ad practices, called Ad Practice Frequency Perception (APFP). We deployed a survey to 1203 members of a representative panel of the US population, and explored whether this metric revealed interesting differentiation across our participants. The survey questions included scenarios in which the stated ad practices were varied across a spectrum from very common to rare or never. Our metric allows us to capture the extent to which users overestimate or underestimate the frequency of a given practice - where users who overestimate believe a practice occurs more frequently than it actually does, and those who underestimate the frequency perceive an ad behavior to be more rare than in practice. Our goal in this initial exploration was not to establish a definitive metric or segmentation, but rather to explore whether this is a promising direction.

We found that the participants' general ability to accurately estimate frequency was fairly poor. Roughly one third underestimate even the common scenarios, while two thirds or more overestimate for practices that are rare. The majority of participants had a fair bit of variance in their answers, implying that they do perceive differences among different scenarios. We used clustering techniques to explore

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

segmenting the population, and our initial results revealed two groups (one large and one small) whose means were consistently and substantially separated across all questions. The smaller group contains people who more regularly underestimate the actual frequency. We also explored some demographic differences between the two groups. Overall, our findings on underestimation and overestimation show interesting variation in user perceptions.

Based on these preliminary findings, we believe perceived frequency warrants further investigation. More work is needed to further segment the larger group since user opinions span a broad range. In the future, this construct might potentially yield an improved ability to characterize and segment users according to their perspective on privacy, particularly in combination with other metrics, although such benefits are as yet speculative.

2. METHODOLOGY

Our methodology is inspired by previous research such as McDonald and Cranor's exploration of people's attitudes and knowledge regarding behavioral advertising, which included questions about the perceived frequency of two scenarios (online ads based on web search history, and online ads based on email content) [13]. Such previous work was largely concerned with establishing people's understanding of practices that were actually currently occurring, often with an eye to informing policy, e.g. [18, 13, 20]. Accordingly, the instruments tended to include one or two questions about how common a practice was and/or whether that practice was allowed by law, combined with a number of other questions such as whether participants felt that practice was beneficial or harmful, their technical understanding of that practice, and their practical behavior related to that practice (such as whether or not they cleared cookies or browsing history) [6, 18, 13, 20].

We build upon and extend this previous work by exploring participants' responses to a larger and wider range of practices, importantly including some that we believe do not currently occur commonly, which allows us to better explore the range of overestimation and underestimation (as opposed to previous work which focused on currently occurring common practices, and therefore naturally primarily explored underestimation). Also in contrast with previous work, we focus explicitly on the use of perceived frequency for characterizing and segmenting users.

Our choice of questions was informed by prior research (e.g. [13, 4]), media coverage, internal studies, and our understanding of common industry practices. Based on these, we generated nine questions about the frequency of a range of potential advertising practices. The questions spanned a variety of services and a variety of entities potentially responsible for advertising or tracking (such as ad companies, email services, insurance companies, photo databases, etc.). The specific questions can be found in the Appendix.[1] A summary of these questions can be found in Table 1, which highlights the type of service each question focused on, and the potential source of advertising in each scenario. For each question, participants were given seven possible answer options based on a 7-point Likert scale (with 1 being the most common, and 7 being the least common). We conducted a round of cognitive testing to refine these questions prior to deploying them.

In addition to these nine questions, we gathered participant demographics, such as age and education level, as well as asking questions about use of the Internet and email to identify non-Internet users.

The questions were administered to 1203 participants in March 2015 as part of a larger survey on privacy attitudes and behaviors. Participants were members of GfK's KnowledgePanel, an online panel based on a representative sample of the United States population (although note that the analysis here is done on individual responses and has not been weighted). The question order was randomized for each participant. For easier understanding, we order the questions in Table 1 based on the trends we observed in the results. We removed data from 56 participants who declined to answer one or more questions and from 55 participants who straight-lined by selecting the same option for all the questions, resulting in a sample size of 1092 participants.

We asked four experts on industry advertising practices to provide responses to the nine questions. One self-declared as a non-expert on some of the questions, and therefore we exclude that person's responses from the analyses. There was nearly complete agreement among the remaining three experts, and their responses were at most 1 point different on the Likert scale for all questions. We selected the median expert response as the *expert answer* in our analysis.

For each of the nine questions, we compare each participant's answer (termed *raw answer*) to the expert answer. We propose the use of an *ad practice frequency perception metric* (APFP) that is defined to be the expert answer minus the raw answer. This metric will then vary between -6 and +6. A value of 0 indicates that the participant gave the same answer as the expert, a negative value indicates that the participant underestimated the frequency of the practice, and a positive value indicates an overestimation. To allow some leniency because it is a somewhat subjective scale, we consider that a participant gave a *correct answer* if the difference is between -1 and +1. Note that while the difference value has a range of +6 to -6, the position of the expert answer on the scale limits the actual allowed range. For example, if the expert answer is 7, the difference values can only range between 0 and +6 (and underestimation is not possible); similarly if the expert answer is 1, then the difference values can range between -6 to 0 (and overestimation is not possible).

3. INITIAL OBSERVATIONS

We first looked to see if any of the participants answered all of the questions correctly, counting a raw answer as correct if it was within +/-1 of the expert answer. Interestingly we found that just 1 of our participants answered all nine questions correctly, just 3 participants answered 8 or more questions correctly, 2.7% answered 7 or more correctly, 11.8% answered 6 or more correctly, and only 42.4% answered 5 or more correctly. This suggests that most people have a somewhat limited understanding of the frequency of current advertising practices.

In Figure 1 we plot the frequency of responses to the questions as a histogram with our APFP metric on the x-

[1]The alert reader may note that AP7, which is inspired by [4], is not specifically about advertising practices, but can readily be extended to these. In hindsight we should have made this revision prior to deployment, and intend to remedy this in future rounds.

Question	App or Service	Potential Advertiser	Expert Answer
AP1	Email	Email provider	Very common
AP2	Shopping	Ad company	Very common
AP3	News	Data broker	Somewhat common
AP4	News	Ad company	Somewhat common
AP5	Online forms	Car manufacturer	Neither common nor rare
AP6	Social	Marketing company	Very rare
AP7	Photos	Photo database	This never happens
AP8	Medical	Insurance company	This never happens
AP9	Restaurants	Search engine	This never happens

Table 1: Ad Practice Question Types

axis. We first look at questions AP1 and AP2, for which the expert answer was "Very common" (a value of 2). Recall that 1 is the highest valued answer a participant can give, and that we consider participants to have answered correctly when they are between -1 and +1 of the expert answer. Accordingly, it is not possible by definition for participants to overestimate for these questions. For these very common scenarios, we observe that the majority of participants estimate these questions correctly in that they are aware that these advertising scenarios are quite common. The rest, about 31-37% of the participants, underestimate the frequency of the practices described in these scenarios. These questions focus on scenarios that involve advertising companies, a data broker and an email provider.

In questions AP3 and AP4, which relate to scenarios that are "Somewhat common" (a value of 3), we observe that a significantly high fraction of participants answered them correctly (85%). Only 8-10% of participants underestimate the frequency of the practices, and a still smaller percentage overestimate the frequency. Interestingly, both these scenarios relate to news services but the potential advertiser is different in each (data broker and advertising company).

For AP5, which relates to the use of contact information entered online being used in telephone marketing, we see that 38% overestimate, 52% get it right, and 10% underestimate. We postulate that users are unaware of where phone call marketing comes from, and may believe that some of it comes from their online activities.

The scenarios in questions AP6, AP7, AP8 and AP9 describe practices that generally do not occur today, such as a health insurance company selling an individual's medical data to an advertising company. We see in the histograms that many participants overestimate the frequency of these scenarios by a substantial amount. 64-95% of participants overestimate these questions, and the amount of overestimation depends upon the question. Even though it is not possible to underestimate for these questions, due to the nature of our scale, we were nonetheless surprised to observe the extent of overestimation. For questions AP6 (social) and AP7 (photo database), the majority of participants significantly overestimate by 2, 3 or 4 Likert scale points; for questions AP8 (medical data) and AP9 (restaurant service/search engine), our participants dramatically overestimate by 4 or 5 Likert points. We might postulate about the source of these misunderstandings: in AP6 it is possible that a fair amount of confusion exists about when such practices do or don't occur in social networks, and this confusion may be bolstered by what people read in the press [7, 11]; in the case of AP7 people might be exaggerating due to press articles about de-anonymization using photos (such as [15] for example); for AP8 people might not realize that different rules apply to medical data; and in the case of AP9 people might be confused about the interaction between search engines and online forms. Clearly these hypotheses need to be explored in further work.

We computed Cronbach's alpha for participants' responses to our 9 questions and found it to be 0.79 which indicates an acceptable level of internal reliability for our exploratory research. We now examine the answers on an individual basis, by plotting the mean and standard deviation for each participant in Figure 2. The black line indicates the mean score for each participant over all nine questions, and the top (bottom) of the vertical line indicates the mean plus (minus) one standard deviation (respectively). The significance of the color is explained later in Section 4. The participants are ordered according to increasing mean. First we observe that there are some users at the far left of this plot who always choose answer options 1, 2 or 3. We found that a total of 103 participants, or 9.4%, did this. These people believe that all of our scenarios are common. While they have correct answers for our 4 common scenarios, they have incorrect answers for the 5 scenarios that are neither common nor rare, are rare, or do not occur. These people can be thought of as *extreme overestimators* as they do not differentiate among the various scenarios and therefore significantly overestimate all of the practices that are rare or do not occur at all.

At the other end (far right) we see only 5 participants who always choose options 5, 6,and 7. These participants consistently select answers that correspond to low perceived frequency of all practices. These people may be thought of as *extreme underestimators* as they seem to believe none of these advertising practices commonly occur.

The vast majority of the participants have a broad range of responses, meaning that they select answers across the range from rare to common. This indicates that participants perceive that ad practices vary across different services and applications, though their perception might be influenced by their lack of knowledge about current advertising practices. Figure 2 also shows that the participants' mean answers vary across a very broad scale, from 1.11 to 6.78, with the majority of the participants' mean answers spread across the range of 2 to 5.

Our survey intentionally included scenarios that range from rare to common, in order to explore both underestimation and overestimation. Overall, we were surprised to see the extent of overestimation. As expected given the nature of the metric, this overestimation occurs predominantly for the scenarios that are more rare.

4. CLUSTERING

We wanted to evaluate if it was possible to group participants based on their answers to the nine questions, and whether these participant groups exhibited different behav-

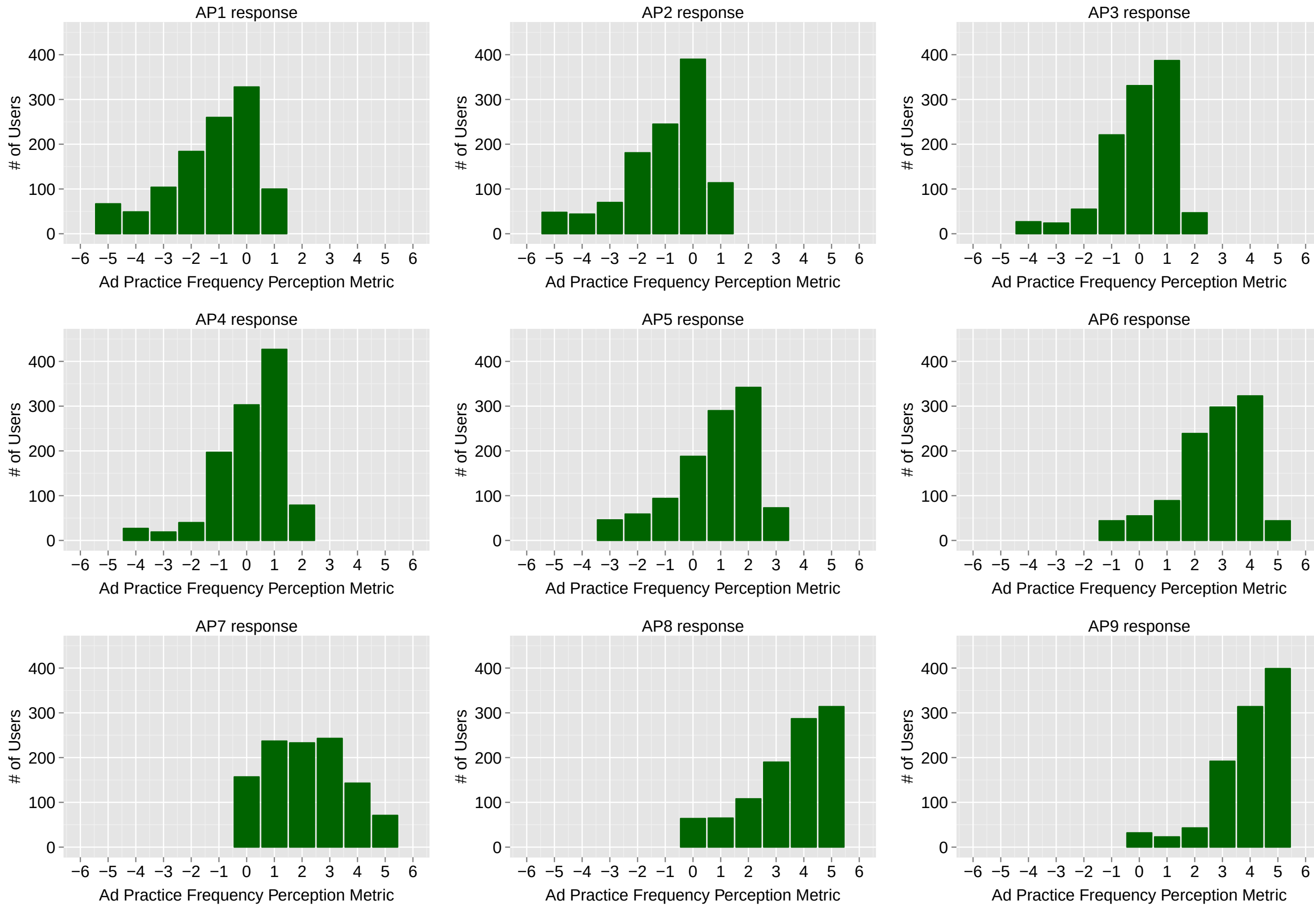

Figure 1: Participant Responses to Ad Practice Questions

iors. Each participant is represented as a vector of 9 numbers, where each number corresponds to the APFP metric calculated for a question based on the participant's answer to the question. These vectors are passed as the input to machine learning clustering techniques.

We experimented with different clustering approaches, such as hierarchical clustering [12] and k-means clustering. In hierarchical clustering we use different agglomeration techniques, such as the Ward method [21], Complete linkage [14], Average linkage [14], and Centroid linkage [14] to merge nearby clusters. While the Ward method minimizes the total within-cluster variance when merging, the three linkage methods merge the closest clusters together and differ in how the inter-cluster distance is calculated. For calculating the distance between participant vectors, we use Euclidean distance, Manhattan distance, and correlation distance.

We experimented with various ways of combining the clustering techniques, 3 distance metrics, and 4 agglomeration methods, and varied the number of clusters from 2 to 10. For comparing the performance of different clustering outputs, we relied on the inherent cluster properties such as the Dunn index [5], Silhoutte width [16], and connectivity [8]. These properties measure the cluster homogeneity (using intra-cluster variance), degree of separation between clusters, and connectedness within a cluster. All the experimentation was done using the R statistical tool and its packages – *clValid* [1] and *hclust* [2].

Based on the three cluster properties, hierarchical clustering using the complete linkage method and Euclidean distance yielded the best results. The best grouping suggests organizing the participants into 2 clusters. The visually interpretable output of the hierarchical clustering is shown in Figure 3. We can clearly see the existence of two user groups (one small with 96 participants and the other large with 996 participants). Visually, we do see the possibility of having more than 2 clusters, however our initial exploration did not result in any clear division among participants in the larger cluster. In future, we plan to explore further if it is possible to regroup the users in the larger cluster into multiple smaller clusters that may exhibit clearly distinct properties.

4.1 Cluster Properties

We observed significant differences among the participants in the two clusters. The scatter plot in Figure 4 indicates the distribution of participants in the two clusters based on their APFP scores computed across the 9 questions. Each point represents an individual participant, where the shape and color of the point are chosen based on the cluster. The points have been jittered (a little random noise is added to the data) in order to see the participant cloud more clearly. The lines indicate the mean ad practice frequency percep-

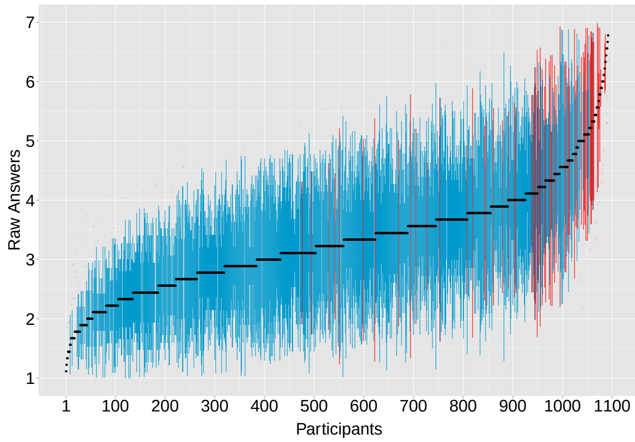

Figure 2: Participants' Raw Answers - Mean and Standard Deviation. Color Corresponds to the Assigned Cluster.

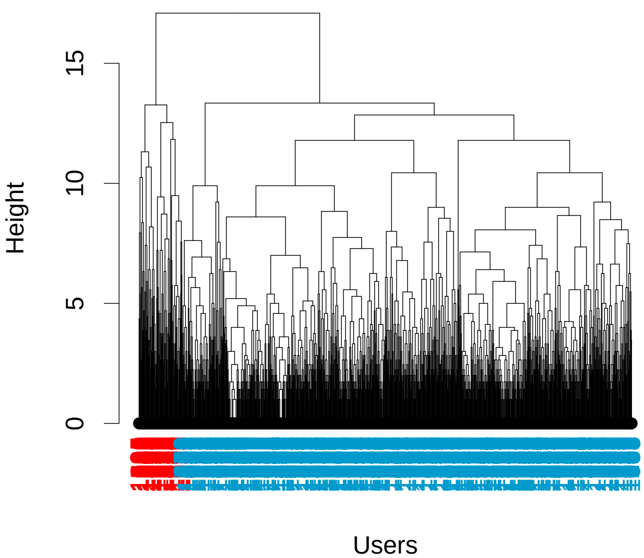

Figure 3: Hierarchical Cluster Dendrogram

tion values computed for each question across participants in each cluster.

As we can see from Figure 4, the mean APFP values for the questions across participants in the two clusters are consistently and substantially separated. The gap between the means across the two clusters varies between 1.0 to 2.36 depending upon the question. The overall mean for Cluster 2 is 4.84 (which rounds to the answer 'Somewhat rare'), whereas for Cluster 1 it is 3.16 (which rounds to the answer 'Somewhat common'). Thus the average difference is 1.68 Likert scale points (on a 7-point scale) between the average participants in the two clusters, a fairly substantial difference which indicates the two clusters do indeed separate participants. This difference between the clusters is also visible in Figure 5, where we show the mean (dots) as well as the mean plus (minus) one standard deviation (top and bottom bar lines, respectively) for the raw answers computed across participants within a cluster. Based on both of these results, we can state that Cluster 2 participants underestimate more

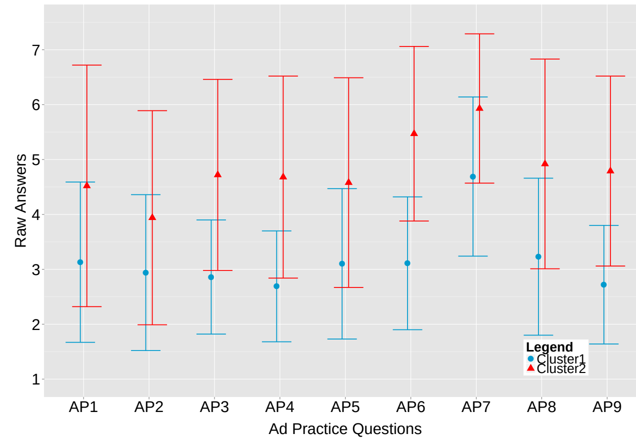

Figure 5: Raw Answers - Mean and Standard Deviation across Clusters

than those in Cluster 1, especially for common ad practices.

Figure 5 indicates that for questions AP3, AP4, AP6, and AP9, there is less overlap between the standard deviation bars across the two participant clusters and that there is consistent difference between the means. This implies that these four questions are possible differentiators of participants across the clusters. It appears that what our clustering algorithm is doing is segmenting out the underestimaters (Cluster 2) largely based upon their answers to questions AP3 and AP4, which represent somewhat common practices that these participants may not be aware of. Also it puts the overestimaters in Cluster 1 largely based upon their exaggerated perception of the frequency of scenarios AP6 and AP9.

Figure 2 shows the participants in the two clusters using different colors. We see a fairly good demarcation between the participants in the two clusters, where the participants in the smaller cluster (Cluster 2) tend to be concentrated near the top right. This also reiterates that the participants in the smaller cluster often tend to underestimate, with their individual means lying around the 'Somewhat rare' and 'Very rare' options.

4.2 Differences in Participant Demographics

To explore if there are any characteristic differences among participants in the two clusters and to see if we can infer some probable causes for the differences in their responses, we studied the demographic details of the participants.

We analyzed the responses of the participants (across the two clusters) to four demographic questions. We noticed no major differences in whether the participants were non-Internet users: 94% of participants in the smaller cluster (Cluster 2) and 97% of participants in the larger cluster (Cluster 1) occasionally used either email or the Internet. Accordingly, Internet usage does not seem to be the primary factor causing the difference in responses across the clusters.

Participants over 50 years old constitute 67% of the smaller cluster and 57% of the larger cluster. So, the fraction of people over 50 years is 10% more in Cluster 2 than Cluster 1. (All participants were over the age of 18, thus shifting up the average age compared with the general US population.) We also observed that the smaller cluster has

Figure 4: Participant Clusters and their Means

10% fewer participants holding a Bachelors or higher degree: 26% in the smaller cluster compared to 36% in the larger cluster. Therefore, Cluster 2 has somewhat older and less educated participants. This may partially explain why Cluster 2 participants underestimate more, although these demographic trends do not appear to have large influence.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we reported the results of an initial exploration of the perceived frequency of various advertising practices, such as those related to email activity, web browsing behavior (search, shopping and news), remarketing, online forms and social networks. We also included scenarios which are rare, or don't happen at all, to examine the limits of people's understanding. Overall participants' ability to accurately estimate which ad practices are frequent or infrequent was fairly poor: only 42% were able to get 5 or more questions right; roughly one third underestimate common scenarios; and anywhere between 64-95% (depending upon the scenario) overestimate scenarios that occur rarely or never. Future work is needed to understand this inaccurate sense of the frequency of advertising business practices.

At the same time, participants did vary their answers across our scenarios, and did appear to realize that multiple practices do exist. We explored a number of clustering methods to see if our population can be naturally segmented into groups of people with similar perceptions. Our best clustering method was able to create two groups with one small cluster (8.8% of participants) clearly distinct from the rest. The difference in average answers between these clusters was 1.68 points on a 7-point Likert scale. We observed that a key difference between these two groups of people lies in their answers to scenarios about news (AP3, AP4), social networking (AP6) and the interaction between search engines and online restaurant forms (AP9). The bulk of the participants in the smaller cluster are *underestimators*. In examining the demographics of these two groups, we found that these underestimators are somewhat older and somewhat less educated than the people in the larger cluster.

Further segmentation of the population in the larger cluster turned out to be challenging as the remaining 91.2% of our participants' opinions span a broad range and they were not immediately separable into consistent sub-groups. Our initial findings in our hierarchical clustering dendrogram indicate that it might be fruitful to explore segmenting the population into as many as 4 to 6 sub-groups. We plan to explore this further in our future work.

Overall, these results suggest that our new construct, based on the perceived frequency of ad practices, together with our preliminary instrument, can reveal people's varying perceptions of the frequency of advertising practices. Our findings on underestimation and overestimation show interesting variation in participants' perceptions. To improve segmentation of the population, we plan to study refining our clusters and making them more granular. We will explore further subdividing our larger cluster into smaller segments by analyzing subsets of questions. We intend to revisit and refine the few survey questions where our experts were not in complete agreement. Importantly, we plan to explore the connection between the perception of online advertising practices with privacy attitudes and behaviors (available from other parts of the larger survey). It is also important to understand how the public's understanding and awareness of these practices evolve over time.

6. ACKNOWLEDGMENTS

We are grateful to Marianne Berkovich and Martin Ortlieb for developing initial versions of three of the scenarios. Our thanks to Mario Callegaro, Sunny Consolvo, Rick Malins, and Lawrence You for helpful advice, input, and support.

7. REFERENCES

[1] clValid: Validation of Clustering Results. http://cran.r-project.org/web/packages/clValid/index.html.

[2] R: Hierarchical Clustering. https://stat.ethz.ch/R-manual/R-patched/library/stats/html/hclust.html.

[3] Workshop on Privacy Personas and Segmentation (PPS). In *Symposium on Usable Privacy and Security (SOUPS 2014)*, 2014.

[4] A. Acquisti, R. Gross, and F. Stutzman. Face recognition and privacy in the age of augmented reality. *Journal of Privacy and Confidentiality*, 6(2):1, 2014.

[5] J. C. Dunn. Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, 4(1), 1974.

[6] S. Fox. Trust and privacy online: Why Americans want to rewrite the rules. In *Pew Internet & American Life Project*, 2000.

[7] V. Goel. How Facebook Sold You Krill Oil. http://www.nytimes.com/2014/08/03/technology/how-facebook-sold-you-krill-oil.html.

[8] J. Handl, J. Knowles, and D. B. Kell. Computational cluster validation in post-genomic data analysis. *Bioinformatics*, 21(15), 2005.

[9] B. P. Knijnenburg. Information disclosure profiles for segmentation and recommendation. In *Symposium on Usable Privacy and Security (SOUPS 2014) Workshop on Privacy Personas and Segmentation (PPS)*, 2014.

[10] P. Kumaraguru and L. F. Cranor. Privacy indexes: A survey of Westin's studies. *ISRI Technical Report*, 2005.

[11] L. Kwoh. Beware: Potential Employers Are Watching You. http://www.wsj.com/articles/SB10000872396390443759504577631410093879278.

[12] C. D. Manning, P. Raghavan, and H. Schütze. Hierarchical clustering. In *Introduction to Information Retrieval*, 2008.

[13] A. McDonald and L. F. Cranor. Beliefs and behaviors: Internet users' understanding of behavioral advertising. In *38th Research Conference on Communication, Information and Internet Policy (Telecommunications Policy Research Conference)*, 2010.

[14] F. Murtagh. A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, 26(4), 1983.

[15] J. Risen and L. Poitras. N.S.A. Collecting Millions of Faces From Web Images. http://www.nytimes.com/2014/06/01/us/nsa-collecting-millions-of-faces-from-web-images.html.

[16] P. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, 20(1), 1987.

[17] S. Spiekermann, J. Grossklags, and B. Berendt. E-privacy in 2nd generation E-commerce: Privacy preferences versus actual behavior. In *Proceedings of the 3rd ACM conference on Electronic Commerce*, 2001.

[18] J. Turow. Americans & online privacy: The system is broken. *Annenberg Public Policy Center, University of Pennsylvania*, 2003.

[19] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang. Smart, useful, scary, creepy: Perceptions of online behavioral advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, pages 4:1–4:15, New York, NY, USA, 2012. ACM.

[20] J. M. Urban and C. J. Hoofnagle. The privacy pragmatic as privacy vulnerable. In *Symposium on Usable Privacy and Security (SOUPS 2014) Workshop on Privacy Personas and Segmentation (PPS)*, 2014.

[21] J. H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 1963.

[22] A. Woodruff, V. Pihur, S. Consolvo, L. Schmidt, L. Brandimarte, and A. Acquisti. Would a privacy fundamentalist sell their DNA for $1000 ... if nothing bad happened as a result? The Westin categories, behavioral intentions, and consequences. In *10th Symposium on Usable Privacy and Security (SOUPS)*, 2014.

APPENDIX

A. AD PRACTICE FREQUENCY PERCEPTION QUESTIONS

For all of our questions, the participants were given 7 possible answers to choose from. We repeat these only once here at the beginning as they are the same for all nine questions.

1. This happens every time
2. Very common
3. Somewhat common
4. Neither common nor rare
5. Somewhat rare
6. Very rare
7. This never happens

AP1. Suppose an Internet user is reading their email online, and their email provider displays ads to them.

In that situation, how common is the following:
The ads are based on the content of email the user sends and receives.

AP2. Suppose an Internet user places a pair of shoes in a shopping cart at a retail website, but doesn't proceed to purchase them.

In that situation, how common is the following:
An ad company associated with the retail website plus other websites tracks this information, and shows the user ads for the same shoes when they visit other websites.

AP3. Suppose an Internet user visits the website for a major newspaper and sees an ad for a hotel chain.

In that situation, how common is the following:
The ad for the hotel chain does not come to the user directly from the hotel chain. Instead, the newspaper buys information from other companies that track and collect information

about which websites individual users visit, and the newspaper uses this information to determine which ads to show to each user.

AP4. Suppose an Internet user visits the website for a major newspaper and sees an ad for an airline.

In that situation, how common is the following:
The ad for the airline does not come to the user directly from the airline. Instead, an ad company determines which ads to show to the user, personally, based on the history of websites they have visited previously.

AP5. Suppose an Internet user is in the early stages of car shopping and visits websites for several car companies. During their exploration, they register at the website of a major car manufacturer and configure a car.

In that situation, how common is the following:
Based on the contact information the user enters at the website, the local car dealership for that manufacturer phones the user to ask if they are interested in test driving a car.

AP6. Suppose an Internet user publicly posts pictures from a recent vacation on a major social network.

In that situation, how common is the following:
A marketing company scans the content of the social network and sells information about the user's interests to other companies, including travel agencies. Later that week they get an email message from an airline company about getaways to similar vacation locations.

AP7. Suppose a person is walking down the street.

In that situation, how common is the following:
A stranger snaps a photo of that person and uses an app on their phone to get information about them. By using online photos and databases, the app determines their name, home address, age, bank account balance, and travel history, and displays that information to the stranger.

AP8. Suppose an Internet user wants to apply for health insurance and fills out an application form on the website of a major medical insurance company, including the fact that they have recently been diagnosed with diabetes.

In that situation, how common is the following:
The medical insurance company sells this information to an online ad company, and the Internet user starts seeing ads about diabetes medication on websites as they surf the Internet.

AP9. Suppose an Internet user searches for restaurant reservations on a search engine, and then clicks through to a restaurant reservation website that is not affiliated with the search engine.

In that situation, how common is the following:
The search engine tracks the information the user enters at the restaurant reservation website, and uses it to customize ads the next time they use the search engine.