Profile CBC: Using Conjoint Analysis for Consumer Profiles

Chris Chapman, Kate Krontiris, John S. Webb
Google

Abstract

We investigate the usage of choice-based conjoint analysis (CBC) for sizing consumer profiles for a technology product area. Traditionally, technology research has often relied upon qualitative personas approaches that are difficult to assess quantitatively. We demonstrate that Profile CBC is able to find consumer profiles from tradeoffs of attributes derived from qualitative research, and yields replicable, specifically sized groups that are well-differentiated on both intra-method and extra-method variables. Thus, we conclude that Profile CBC is a potentially useful addition to analysts' tools for investigating consumer profiles.

Introduction: The Business Problem: Sizing Consumer Profiles

The Google Social Impact team works on products and technical ecosystems for social good. This includes work on crisis response, civic innovation, and other social areas. For the project here, the team was interested to enhance civic engagement. As an example of the products this might inform, consider information served to users in advance of the November 2014 U.S. midterm election. The Civic Innovation team proactively served election information to Google Now users with four information designs; two such designs are shown in Figure 1.
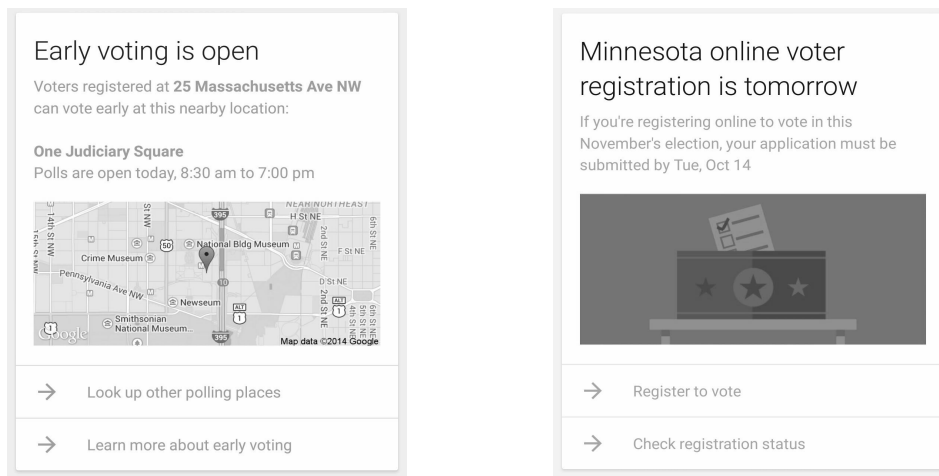

Figure 1. Example Information Cards from Google Social Impact, October 2014

Serving these cards assumes that many users will find the information useful even though they might not have sought it. Previous qualitative research characterized such users as

"interested bystanders," people interested in civic life yet who are not necessarily active participants or seekers of information (Krontiris et al, 2015).

Krontiris et al (2015) documented civic *personas*, descriptions of prototypical (not actual) users. Personas are commonly used in technical product development to build product team awareness of users and to inspire design solutions. Personas may compile personal, behavioral, motivational, and product interest characteristics. An example excerpt is shown in Figure 2.

> *Kathleen is 33 years old and lives in Seattle. She's a stay-at-home mom with two children: Katie, 7, and Andrew, 4. She drives the kids to school (usually carpooling with 2-3 other kids) in her Volvo wagon. Kathleen is thinking about buying the Sony rear-seat entertainment system she saw last weekend at Best Buy to keep the children occupied on the upcoming trip to see family in Canada.*

Figure 2. Excerpt from an Example Persona (Brechin, 2008)

Before committing to projects that push civic information to users, the Google team wished to know how many people would benefit. Thus, the key business question was, "*How many interested bystanders are there [in the United States]*?" In other words, how many people might benefit from Google Now cards that proactively present information about civic events?

Difficulty with the Business Question

Unfortunately, as a qualitative description of a prototypical customer, a persona is not immediately sizable. In the present project, the qualitative research provided descriptions of purported representative interested bystanders but did not specify how many there were. This situation reflects two problems for personas: that, as pure descriptions, they are neither confirmable nor falsifiable (Chapman & Milham, 2006); and that, as composites of multiple dimensions, they fall prey to the curse of dimensionality. Once a persona comprises more than a few attributes, it is likely to match no one in an actual population (Chapman et al, 2008).

For these reasons, the first author had typically advised business stakeholders not to use qualitative personas in efforts to do market sizing. Instead, he has suggested that personas should be viewed as inspirational rather than descriptive. In this paper, however, we propose that choice-based conjoint analysis offers an appealing alternative that allows integration of multiple qualitative attributes while allowing quantitative sizing of groups.

Method: Choice-Based Conjoint Analysis Profiles, or Profile CBC

We addressed the problem of sizing the civic profiles using choice-based conjoint analysis (CBC), where the attributes were not product characteristics but were instead attitudinal

and behavioral statements characteristic of persona attributes. The attitudes were derived from consumer characteristics that had been observed in the preceding qualitative research.

These characteristics were arranged into common areas (CBC attributes) comprising statements that could be considered to trade off against one another (hence, attribute levels). The list of characteristics included 8 areas (attributes) with 3-4 statements in each area (levels), for a total of 27 levels. Selected CBC attributes and example levels are shown in Figure 3.

Civic engagement	4 levels: I don't have time …; I try to do as much as I can …; etc.
Family engagement	3 levels: I don't spend very much time with my family; etc.
Career engagement	3 levels: My career or education is my main priority …; etc.
Attribute D	3 levels
Attribute E	3 levels
Attribute F	3 levels
Attribute G	4 levels
Attribute H	4 levels

Figure 3. CBC Attributes and example levels. Attributes D-H are disguised in this paper.

This CBC design was fielded as a *partial profile CBC* (Chrzan and Elrod, 1995) such that each task presented *three concepts (profiles)*, where each profile comprised levels from *three* of the eight attributes. As we will describe below, we found this CBC format to be optimal for respondents' ability to perform the task. Also, as will be explained below, there was no "none" response option. An example task as fielded is shown in Figure 4.

Thinking about civic or community engagement, which one of these PROFILES is more like YOU?

Choose *which profile is more like you* by clicking one of the buttons below:

	Profile 1	**Profile 2**	**Profile 3**
Career Involvement	I'm not working or in school right now.	My career or education is my main priority right now.	I balance my career or education with other obligations and pursuits.
Civic engagement (volunteering or community activity)	When I have free time, I spend it on civic or community activities.	I don't have time for civic or community activities.	I try to do as much civic engagement as I can, but I have other obligations.
Family involvement	I balance family time with career and social pursuits.	I don't spend very much time with my family.	I spend as much time with my family as I can.
	○	○	○

Figure 4. An example partial profile CBC task, as fielded.

Each respondent answered 12 tasks (with 3 profiles each). The attributes shown were randomly selected and ordered from all eight attributes and varied from task to task. The survey fielded a total of 500 variations of the 12-task questionnaire, created with the Sawtooth Software

SSI/Web CBC module, and each respondent was randomly assigned to one of the 500 variants. Respondents were adults in the United States, obtained through an internet panel fielded by a third party market research supplier in October 2014. The data comprised N=2087 complete responses to the survey.

After data was collected, we identified profiles using aggregate latent class analysis of the conjoint utilities, conducted with Sawtooth Software CBC Latent Class (Sawtooth Software, 2004). As we discuss below, an alternative would have been to perform market simulation for specified profiles.

How Conjoint Analysis Solves the Sizing Problem

For experienced conjoint analysis analysts, the answer may be obvious: conjoint analysis provides utility estimates that allow determination of a probability estimate for each individual's match to a particular set of attributes (i.e., profile or persona), compared to other sets.

For analysts who are new to conjoint analysis, this works briefly as follows. For each respondent, a statistical model estimates a metric part-worth "utility" reflecting the preference for each of the attribute levels. The partworths or utilities reflect the best estimate for likelihood to prefer one choice (in this case, one profile) in comparison to a specified set of alternative profiles, with likelihood proportional to the share of exponentiated summed utilities in the multinomial logit (MNL) model.

For illustration, consider utility values obtained for 2 levels (1 and 2) of two attributes (A and B), where each attribute level represents a statement as in Figures 3 and 4. Suppose for one respondent, utility(A1) = 0.5, utility(A2) = -0.4, utility(B1) = 1.1, and utility (B2) = 0.0. Suppose further, we are interested to compare two profiles for this respondent: profile 1 that comprises levels A1 & B2, and profile 2 that comprises A2 & B1.

Under MNL, the proportion of preference for each profile is expressed as share = *exp(*for one profile *x,* sum*(utilities(profile x))) /* sum*(for all profiles i, exp(*sum*(utilities(profile i)))*. In the present case, sum*(utilities(profile 1)) = 0.5 + 0.0 = 0.5*, and sum*(utilities(profile 2)) = -0.4 + 1.1 = 0.7*. This gives exponentiated values for profile 1 = *e^0.5 = 1.64* and profile 2 = *e^0.7 = 2.01*. Taking the share of preference ratios, the likelihood that this respondent matches profile 1 better than profile 2 is calculated as *1.64 / (1.64 + 2.01) = 45%*, and similarly the likelihood of better matching profile 2 is calculated as 55%.

Note that such allocation is only defined relatively within a specified set of profiles using various levels drawn from the same attributes; it does not answer the question whether some other, unknown profile might fit better. To identify the most likely profiles, rather than simulating them exhaustively we used latent class analysis to identify groups. In the discussion section below, we consider alternative methods to identify profiles.

Such assessment is based on the respondent's own answers to the profile questions, and takes into account the contribution of each attribute for that respondent. In this way, it solves the

difficulty of allocating respondents to profiles in the face of multidimensionality and imperfect matching. We leave aside details such as how part worth estimates are calculated; for further discussion of MNL, we refer readers to Orme (2009).

Results: Answering the Business Question

Latent class solutions on the conjoint utilities were found for 2-10 classes, and a final solution of 6 classes was selected as preferable according to several criteria. In particular, the 6 class solution showed stronger fit indices (AIC and BIC) than solutions with fewer classes; the classes were qualitatively well differentiated and interpretable; the class sizes were relatively uniform, ranging 11%-23% of the sample; and solutions with more than 6 classes demonstrated weaker fit indices, less interpretable differentiation, or undesirably small groups (e.g., fewer than 5% of respondents in one of the classes). Among multiple versions of a 6 class solution proposed by CBC Latent Class, we retained the solution with the best fit index (AIC and BIC).

An overview of the 6 class solution results for the first three attributes is shown in Figure 5, which shows the mean part worths (utility values) for each level of each attribute, by each group (note that the part worths as shown are rescaled to be comparable and are not on the raw scale suitable for MLM calculation as shown above).

	Part Worth Utilities Rescaled for Comparability						
	Segment Sizes	14.70%	20.70%	15.30%	22.60%	11.00%	15.80%
		Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Career Engagement	My career or education is my main priority right now.	-62.58	29.82	-94.28	-2.83	42.79	55.17
	I balance my career or education with other obligatior	-19.04	88.25	-26.82	-15.64	51.92	89.57
	I'm not working or in school right now.	81.62	-118.07	121.10	18.47	-94.70	-144.74
Civic Engagement	I don't have time for civic or community activities.	-5.19	-47.97	50.04	24.07	4.85	70.16
	I try to do as much civic engagement as I can, but I h	29.41	44.49	0.85	22.91	24.16	-3.76
	When I have free time, I spend it on civic or communi	24.35	20.49	-23.39	-53.94	-19.42	-39.49
	My profession is a form of civic activity.	-48.56	-17.02	-27.49	6.97	-9.60	-26.91
Family	I balance family time with career and social pursuits.	19.01	50.47	10.13	13.98	26.63	38.16
	I spend as much time with my family as I can.	67.35	36.16	73.24	-122.02	35.10	30.01
	I don't spend very much time with my family.	-86.36	-86.62	-83.37	108.04	-61.73	-68.17

Figure 5. Excerpt of part worth means for the 6 class solution. Part worth values are shaded to indicate direction and magnitude, and are not raw utility values but have been rescaled to be comparable.

In Figure 5, we see that the classes are well differentiated from one another across the rows. For example, in the "Career Engagement" attribute, Groups 1 and 3 often chose profiles without work or study, whereas Groups 2, 5, and 6 were likely to work. Additionally, each profile showed some attributes that were strongly loaded on it, within the columns. For instance, Group 4 is heavily identified as not spending time with family, and Groups 3 and 5 identified not having time for civic activities.

In short, the 6 class solution was interpretable, differentiated, and was free of the common but undesirable residual class (a class where no attribute is strongly associated, and the class is uninterpretable).

What about the business question? How many interested bystanders were there? Of the 6 classes, the utilities for 3 classes showed weak engagement in civic activities yet simultaneous high interest in civic happenings and information sources such as news. We identified these as matching the "interested bystander" profile; they correspond to groups 3, 4, and 5 in Figure 5.

Figure 6 presents the six groups with brief descriptive names and sizing. The interested bystander groups are the "Absentees," "Issues-Aware," and "Vocal Opinionator" groups, and comprise an estimated 48.9% of the respondents.

Given this breakdown of the groups' sizes, the business stakeholders concluded that there were enough interested bystanders to warrant further investigation of their needs and product design to meet those needs. Additional detail about civic engagement behaviors from the profiles (not shown here) provided more specific points to address with interested bystanders.
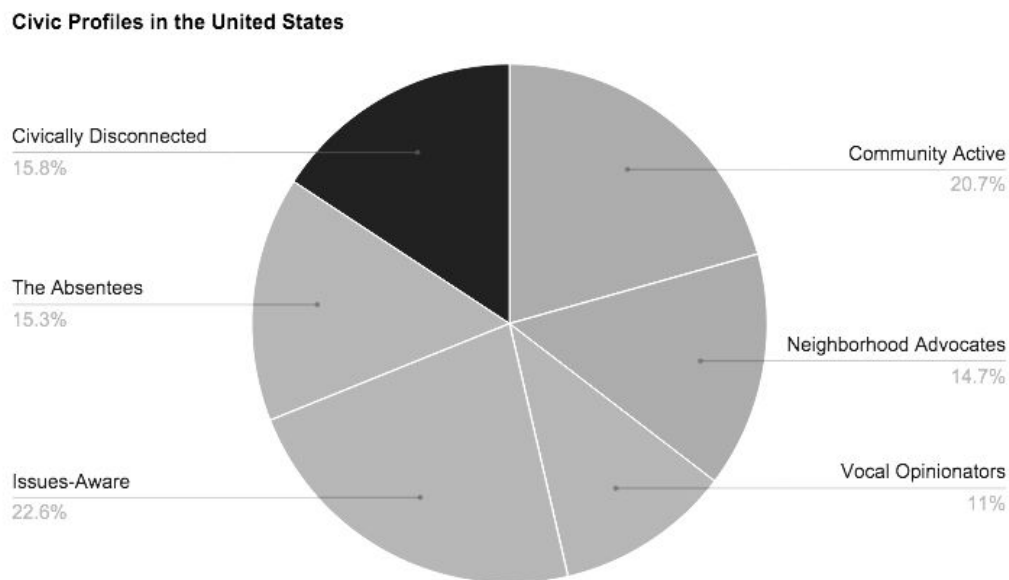

Figure 6. Sizing and descriptive titles for the six identified profiles. Interested bystanders comprise the "Absentees," "Issues-Aware," and "Vocal Opinionator" groups.

External Correlates

A frequent outcome in segmentation projects is that class membership is strongly related to the basis variables used to classify people, in this case the conjoint utilities, yet the groups are weakly or not at all differentiated on other variables. In the present survey, we collected data separately from the CBC exercise on several other variables: household income, work status,

gender, and self-reported frequency of voting. Figure 7 shows the group level means on those external covariates for each of the 6 civic profiles.

	Community Active	Neighborhood Advocates	Vocal Opinionators	Issues Aware	The Absentees	Civically Disconnected	*Range of group means*
Est'd mean income ($)	49,965	71,985	41,511	51,272	76,184	61,943	*34,673*
Employed full time	20%	66%	13%	41%	61%	59%	*53%*
Female proportion	58%	47%	66%	39%	47%	55%	*27%*
Report routine voting	55%	61%	37%	41%	48%	34%	*27%*

Figure 7. Mean by civic profile class for behavioral and demographics measures.

In Figure 7, we see that the six classes are once again well differentiated on the external variables. For example, full time employment ranges from 13% to 66% across the groups for a total 53 point spread from highest to lowest. There is a 27 point spread in gender and 27 point spread in reported voting frequency.

It is important to remember that these variables were not used in the profile determination and respondent assignment, and the clear differentiations here both confirm the importance of the profiles found and their external validity with regards to important civic behaviors. In other words, the Profile CBC method yielded profiles with important differences on other measures.

Discussion: Profile CBC Task Design

The present study reflects several rules of thumb for task design that the authors have formulated in the course of attempting Profile CBC in several product categories with several audiences. We offer these as best practice recommendations with the caveat that they are entirely based on our limited experience; we hope additional research will strengthen or modify them.

Our six suggested design principles are presented in Figure 8.

1. Be careful to omit "must-have" attributes
2. Tasks should consist of 2 or 3 concepts
3. Concepts should present partial profile limited to 3 attributes
4. Tasks should not use a "none" option (especially single response "none")
5. Tasks should not use allocation CBC
6. Careful investigation is needed before using ACBC

Figure 8. Suggested design principles for Profile CBC questionnaires

Principle 1 -- to omit "must have" attributes, means to ensure that all levels are actually suitable for trade-off by respondents. If a level is crucial to someone's self image or choice to the point that he or she would find it impossible to choose a conflicting profile, that level is better omitted (and perhaps could be used as an external validation measure instead). For instance, gender might fall into this category.

Principles 2 and 3 -- that tasks should present 2 to 3 concepts and no more than 3 attributes -- reflect the cognitive difficulty of the task otherwise. With pre-testing, an analyst might cautiously relax these, but 3 profiles and 3 attributes is the maximum that we have found to be comfortable for respondents (cf. Patterson and Chrzan, 2003, for more on how respondents handle partial profile tasks). Principle 4 -- to avoid "none" -- reflects the fact that respondents may use "none" with extreme frequency when selecting profiles if any level is an imperfect match. Because we are interested in their tradeoffs, it is difficult to interpret what "none" would mean in the context of Profile CBC.

Principle 5 -- to avoid allocation CBC -- arises because of the cognitive complexity of "allocating" oneself across multiple profiles. Principle 6 -- to be cautious with ACBC -- arises because of the difficulty in presenting screening tasks in ACBC. We have attempted but so far not succeeded in wording ACBC screening tasks in a way that works for respondents. If this problem were solved, we believe ACBC would have potential for Profile CBC.

Market Simulation: Alternative to Latent Class

In the present project, we began with qualitative personas, extracted their attributes, fielded a conjoint analysis study, and used latent class analysis (LCA) to determine profiles. This effectively discarded the original personas in favor of new ones (although largely similar to the previous personas) as found by LCA.

An alternative process would be not to use LCA, and instead to specify profiles after conjoint analysis whose attributes match those of the qualitative personas. We could then use the multinomial logit share formula or other market simulation techniques to determine the proportion of match for each of the specified profiles. This would be identical to the procedure outlined above in the section, "*How Conjoint Analysis Solves the Sizing Problem.*"

For the present study, we used the LCA profiles instead of market simulation for the qualitative personas for three reasons. First, the LCA results were similar enough overall to the personas that they were able to answer the business question and afforded the advantage of "letting the data speak." Second, LCA addresses gives quantitative guidance as to how many profiles there should be.

Finally, even when care is taken to select attributes, it is difficult to construct market simulations that precisely match a qualitative profile. For instance, suppose we are considering an attribute that is related to a persona but is of comparatively lesser importance than others. Should it be included in the market simulation or not? One could argue either way, and the

choice will affect the share estimates. It cannot simply be determined by running both models because that would end up with cherry picking and steering the outcome. Because this situation may arise for many attributes across multiple profiles, it can lead to uncertainty in how to set up a market simulation.

With those caveats, we feel that market simulation is feasible and worthwhile when profiles are carefully constructed. Market simulation could also be used as a test of specific alternatives. For instance, if one asked, "does this profile fit better than this other one?" it would be straightforward to put both into a market simulator and assess the relative shares of each.

Portfolio Modeling: Alternative to Latent Class Analysis

We used LCA to find the classes here, but there are several alternatives. As noted above, one simple alternative is to specify the classes directly and use a market simulator to size them. One possibility is to apply statistical procedures that have various assumptions other than the LCA methods used here. For instance, one might use Gaussian finite mixture models (cf. Benaglia et al, 2009).

Another alternative is a portfolio modeling approach that attempts to find the best set of profiles to match the respondents, as based on an overall fit criterion such as proportion of respondents matching or maximum likelihood. These may be performed through various iterative search techniques (cf. Chapman and Alford, 2010). For general optimization methods, a crucial question to consider is whether one needs answer the problem of a "none" parameter. If an algorithm simply maximizes total share of people allocated to groups, then it will always "succeed" by allocating 100% to a single group unless there is a criterion such as a none utility that prevents such allocation. However, as noted above, the "none" concept is difficult to express here; this problem -- of how to express "none" and model it -- is an area ripe for investigation.

MaxDiff: An Alternative to CBC

MaxDiff is a potential alternative to choice-based conjoint analysis to field these tradeoffs. In particular, if the attributes on a profile are considered to be a list of characteristics that each might or might not apply and are not necessarily arranged into crisp groups (i.e., attributes), then a MaxDiff approach should work well. Additionally, MaxDiff might be conceptually simpler for respondents. Latent class analysis would work with MaxDiff responses nearly identically to the procedure we described here for Profile CBC.

Additional Notes on Partial Profile Designs

We have argued that partial profile tasks make Profile CBC possible. We believe that fielding a choice task with more than a few attributes that ask about self-identification all at

once, such as the fictional one shown in Figure 9, is simply infeasible. In pre-tests, respondents balked at such a task.

Which of these profiles best matches you?

Profile A	**Profile B**	**Profile C**
Loves golf	Loves football	Doesn't like sports
Very smart	Average smart	Average smart
Not married	Not married	Married
Kids at home	Kids at home	No kids at home
Prefers jazz	Prefers classical	Prefers hip-hop
Hates pizza	Loves pizza	Loves hamburgers
Drives Chevy	Drives Volvo	Drives Ford
Average weight	Underweight	Overweight
Goal = be happy	Goal = money	Goal = balance

Figure 9. An example (fictional) Profile CBC task that avoids partial profile design but may be impossible for respondents.

However, there are potential concerns with partial profile designs. For one, the required sample sizes will increase substantially; instead of a few hundred respondents, more might be required because each task provides less information. We suggest testing the design matrix to ensure there is adequate power for the intended sample size.

Because partial profile designs do not show all attributes in each task, they may underestimate the extent to which attributes are correlated. To see why this is, suppose that attributes A and B are very closely related. In a partial profile design, A and B often will not appear together. Thus, when each is observed without the other, it appears to contribute the full effect by itself in impact on the choice task, and this does not reflect its overlap with the other attribute. Additionally, because A and B often appear separately, there is corresponding less opportunity to assess tradeoffs between their various feature levels. This issue of potential attribute correlation should be examined at design time with respect to theory and previous findings, subjected to pre-testing before fielding a final survey, and examined post hoc for either excessive correlation or absence of expected correlation.

In the present study, we investigated attribute correlation qualitatively before fielding, and empirically after fielding the study. Figure 10 presents the Pearson's *r* correlation matrix for part worth utilities found in this study, where the circle shading indicates direction (positive correlation is lighter and negative correlation is darker) and circle size indicates magnitude (plotting method from Wei, 2013). Overall, we see that several of the attributes are substantially correlated (e.g., attributes B, E, F, G, and H). These correlations were expected on theoretical bases for the attributes in question, and thus the correlations were confirmatory. Likewise, much of Figure 10 shows correlations of low magnitude (small circles) between levels; this was likewise confirmatory for attributes that were expected to have lesser levels of association.

Overall, we conclude that the partial profile method is likely required for Profile CBC, and that the problems of power and attribute correlation may be managed with attention and post hoc empirical inspection. To review more about partial profile concerns, see several papers in previous Sawtooth Software Conference proceedings (e.g., Huber, 2012; Yardley, 2013).
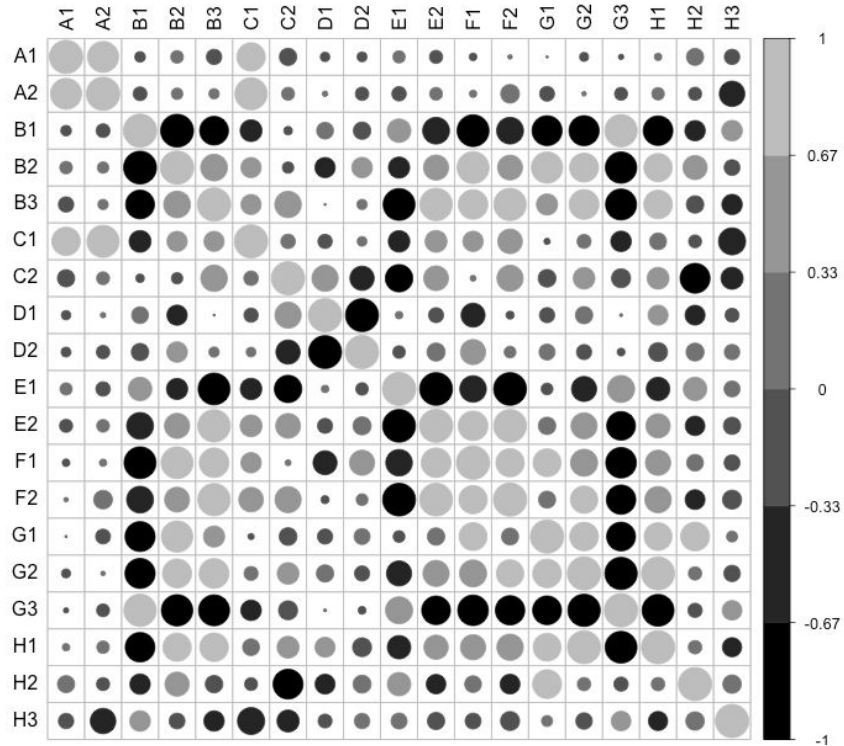

Figure 10. Correlation matrix for the attributes in the present study, N=2087. The final level in each attribute has been omitted. Circle size is proportional to absolute magnitude of correlation, and hue indicates direction.

Conclusion

The Profile CBC method outlined here demonstrates that choice-based conjoint analysis may be useful in situations where analysts seek to find and size clusters of respondents who identify with profile-like descriptions. Because Profile CBC allows incorporation of qualitative self descriptions as attributes, finds classes with a replicable procedure, and determines class size, it overcomes key limitations of purely qualitative personas. In the present study, we also observed that the classes showed substantial discrimination on external validation measures. Thus, when basic design cautions are observed, Profile CBC opens exciting new areas of exploration for conjoint analysts.

References

Benaglia, T., Chauveau, D., Hunter, D.R., Young, D. (2009). mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software*, 32(6), 1-29. http://www.jstatsoft.org/v32/i06/ (Last retrieved May 2, 2015.)

Brechin, E. (2008). Reconciling market segments and personas. Cooper Design. http://goo.gl/btJXsP (Last retrieved May 4, 2015.)

Chapman, C.N., and Alford, J.L. (2010). Product Portfolio Evaluation Using Choice Modeling and Genetic Algorithms. *Proceedings of the 15th Sawtooth Software Conference*, Newport Beach, CA, October 2010. https://goo.gl/fM86kp (Last retrieved May 4, 2015.)

Chapman, C.N., Milham, R.P. (2006) The personas' new clothes: methodological and practical arguments against a popular method. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50:5, 634-636. SAGE Publications. https://goo.gl/szQ54E (Last retrieved May 4, 2015.)

Chapman, C.N., Love, E., Milham, R.P., ElRif, P., and Alford, J.L. (2008). Quantitative evaluation of personas as information. *Proceedings of the Human Factors and Ergonomics Society (HFES) 52nd Annual Conference*, New York, NY, September 2008. http://goo.gl/4rLYEO (Last retrieved May 4, 2015.)

Chrzan, K., and Elrod, T. (1995) "Partial Profile Choice Experiments: A Choice-Based Approach for Handling Large Numbers of Attributes." Presented at the 1995 Advanced Research Techniques Forum, Monterey, CA.

Huber, J. (2012) CBC Design for Practitioners: What Matters Most. *Proceedings of the 16th Sawtooth Software Conference*, Orlando, FL, March 2012. http://goo.gl/ieqgMK (Last retrieved May 4, 2015.)

Krontiris, K., Webb, J., Krontiris, C., Chapman, C. (2015). Understanding America's "Interested Bystander:" A Complicated Relationship with Civic Duty. Technical report, Google Civics Research Workshop, New York, NY, January 2015.

Orme, B.K. (2009). *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*, 2nd edition. Madison, WI: Research Publishers.

Patterson, M., and Chrzan, K. (2003). Partial Profile Discrete Choice: What's the Optimal Number of Attributes. *Proceedings of the 10th Sawtooth Software Conference*, San Antonio, TX, April 2003. https://goo.gl/uNqjTt (Last retrieved May 4, 2015.)

Sawtooth Software (2004). The CBC Latent Class Technical Paper (Version 3). Sawtooth Software, Sequim, WA. http://goo.gl/n2ZmAu (Last retrieved May 4, 2015.)

Wei, T. (2013). corrplot: Visualization of a correlation matrix. R package version 0.73. http://CRAN.R-project.org/package=corrplot (Last retrieved May 4, 2015.)

Yardley, D. (2013). Attribute Non-Attendance in Discrete Choice Experiments. *Proceedings of the 17th Sawtooth Software Conference*, Dana Point, CA, October 2013. http://goo.gl/2Yg91Z (Last retrieved May 4, 2015.)