Research Article

Steps towards an Integrated Database of the Citizen's Mind Using Mind Genomics

Howard Moskowitz[1]*, Arthur Kover[2], Eric Grunfeld[3], Hollis Belger[4], Laura Estefania Rodriguez[5] and Joshua Deitel[6]

[1]Mind Genomics Associates, Inc., White Plains, NY, USA

[2]Retired Faculty, Fordham University, New York, USA

[3]Brown University, Providence, Rhode Island, USA

[4]Independent Researcher, California, USA

[5]Independent Researcher, Bogota, Colombia

[6]Sifra Digital, Jerusalem, Israel

*Corresponding author: Dr. Howard Moskowitz, Mind Genomics Associates, Inc., White Plains, NY, USA; **Email:** mjihrm@gmail.com

Abstract

We present an approach to database the mind of the citizen on topics, using a set of interlocked studies created through Mind Genomics, in which the elements stay the same, but the topic changes. The database allows the creation of models, equations relating the rating of systematically varied vignette to the presence/absence of 16 elements (statements), as well as estimating the effects of features describing the respondent (gender, age, belief in what method best solves social problems). The data uses experimental design to the create the test stimuli (vignettes), dummy variable regression analysis to show the contributions of the elements and the features describing the respondent, as well as clustering to create new to the world mind-sets, different ways to look at a topic. The paper closes with the suggestion of how to create these databases in either an ad hoc fashion, or preferably in a systematized way, year over year.

Introduction

The academic dealing with social issues, especially problems dealing with the plethora of economics-based problems, is simply enormous, and cannot be straightforwardly summarized. What is missing, however, appears to be an integrated approach to studying these social issues from the mind of the typical citizen. There are various public polls of consumer sentiment), and one-off polls, really studies, about current issues, usually sponsored by an organization involved in public affairs and conducted by a market research company using strict rules of consumer research (e.g., Axios polls conducted by IPSOS, a marketing research conglomerate well known to fits work in the area).

The information about public issues, e.g., studies about what bothers people, appears in documents, summarized, and simplified for public consumption by the media. The rest of the information may go into the innumerable topic-related books published by commercial publishers, or go into reports circulated to politicians and other public servants. Studies such as the Quinnipiac polls [1] are executed year after year, and the database compiled, both for those interested in current problems as well as those interested in the study of changing social scene over time. Some of the thinking can be traced to the discipline known as SSM, soft systems methodology [2].

For studies done an ad hoc basis, there is no reason to create this integrated database of the mind. Such as concept might be interesting, but it does not fit the view of those who want to focus on the moment, and report what is happening in the 'her and now.' Those who database their information would be more likely to appreciate an integrated database, contents able to be cross-referenced. Classic books on the mind of the citizen in society might have benefited from the availability of such a database, although that statement is more of a conjecture than a point of fact. Yet, we might consider how earlier efforts might have been enhanced by this type of data, such as the pioneering book by [3]. The current precis of their book, available in 2022, describes the research effort for which the database might be invaluable:

Presents a review and analysis of theoretical and empirical issues in the mechanisms and functions of interpersonal behaviors and their development in social encounters. The relationship of social cognitive structures in the individual to societal structures, developmental, emotional, and economic aspects of interpersonal relations... [4]

The vision of Mind Genomics to provide a database of the citizen's mind began in the early 2000's. At that time there was a growing interest in the mind of the citizen about social issues. Twenty years, ago, however, the focus was simply on understanding social issues from the inside of a person's mind. The senior author participated in studies of response to the voting platform of candidates e.g., the voting platform of Kerry [5]. Inspiration for the work came from the newly emerging interest in computers for data acquisition and the use of experimental designs to create combinations of ideas that the respondent would then evaluate [6]. The effort continues today, suggesting that there is an underlying current of acceptance of

conjoint measurement to understand citizen minds [7], along with the recognition that understanding mind-sets can help impact education for students, and create a better world [8].

At the same time, there was an obvious lack of integrate databases about the mind of the citizen about ordinary problems of daily life. The media as well as the journals were and continue to be populated with either continuing stories in the case of media, or well executed but one-off studies by academics using the most powerful social science tools. The senior author executed one large scale study on different situations causing anxiety, using the Mind Genomics tool described below, finding the approach to generate a reasonable integrated database. That database revealed far more than would have been revealed by 15 disconnected studies on the same topics. The success of integrated the 15 parallel studies into a single database called 'Deal With It' (for colloquiality) generated the vision that one could use the disciplined approach by the emerging science of Mind Genomics to create a database of the citizen mind, and perhaps make a contribution to the emerging discipline of citizen science [9-11].

The Mind Genomics Approach and Its Use in a Societal Issues Database

Key issues facing the citizen are often approached by researchers using qualitative (depth) interviews, either with single individuals or groups, usually to get a sense of 'what's happening in the mind of the citizen.' Beyond that there may be polls or surveys about the topic. Beyond that is the sociological approach of looking at people in groups, as well as studies of the way a society works. There are no databases to speak of which go into the mind of the citizen, at least no systematized databases updated on a yearly basis, across aspects of the citizen's life.

Traditional research answers the various questions in an adequate way, but often the data is in a somewhat disorganized format because there is the need to tell a coherent story after digesting and integrating the various sources and types of information. The astute, insightful investigator can pick up the thread of the story, and, with the right data, weave the story together so it morphs into a compelling narrative. When the topic is of sufficient importance, other efforts may be initiated to fill the gaps, and round out the topic.

What is missing from the foregoing is a systematic way to explore the world of the citizen from the inside of the citizen's mind, doing so with groups of related topics, doing so with people around the world, and on a systematic basis. The data produced by a systematic approach can become invaluable, supplying insights, revealing patterns, increasing our factual knowledge, and promoting the discovery of patterns. If, perchance, the approach is also affordable, then society has the capability to profile itself, worldwide, over time, creasing a database that might well reveal short term and long term patterns.

Mind Genomics as an Affordable, Efficient, Scalable System

The entire Mind Genomics process is templated, from start to finish, including the analysis. Through the templating, the technology forces the researcher to learn a new way of disciplined thinking, a way which ends up being an algorithm for solving a problem, or even for innovation. We begin with the three steps, shown in Figure 1.

Step 1: Choose the Topic

Figure 1 (left panel) shows that the study topic is 'Loss of Hope.' The database includes the results from these five economics-oriented studies, chosen from the full set of 26:

1. College Expense - Education for people in College is too expensive.

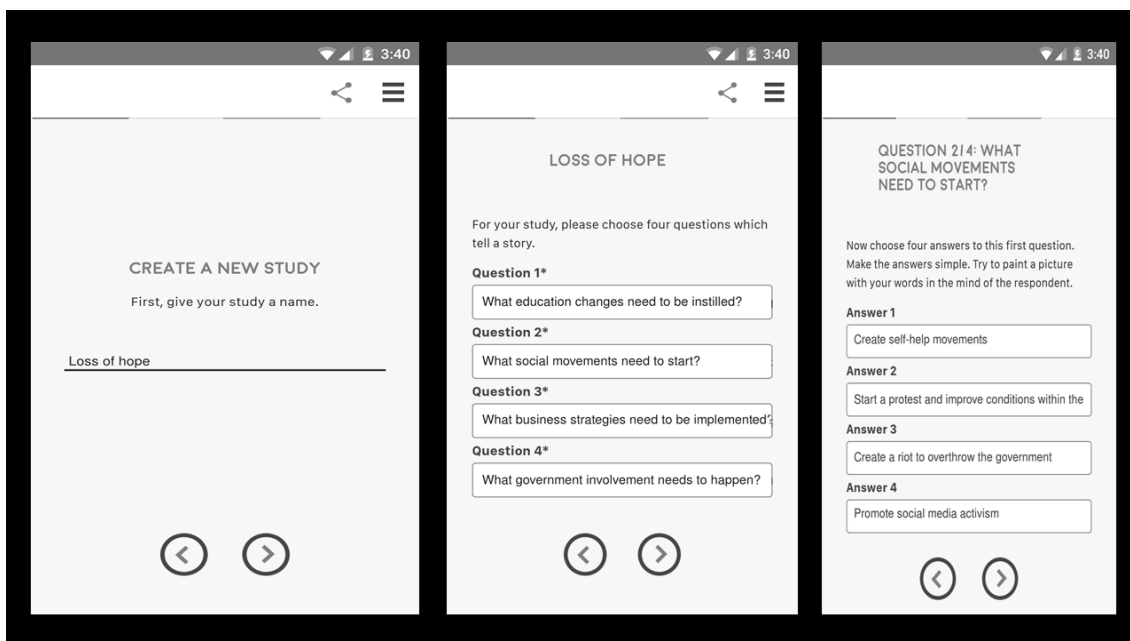2. Economic Gap - Rich people get richer, everyone else falls behind.

Figure 1: The first three templated steps for set up, choose the topic (left panel), select four questions (middle panel), and generate four answers to each question (question 2 on the right panel).

3. Loss of hope - People who have no hope that anything they do will help their lives.

4. Poverty – Poverty so that some people don't have enough to eat.

5. Social Security – People not sure that Social Security will last.

Step 2: Select Four Questions or Dimensions Which 'Tell a Story'

Our 'story' is not a story but rather four sources of solutions (education, social, business, and governments.

Step 3: Create the Element, Four Specifics from Each Type of Solution, or 16 Elements

The set of 26 studies dealt with the solutions of social problems. The solutions were to be appropriate to 'solving' the fundamental or underlying issues which led to the problems, not to the actual specific solutions, which would be topic-specific, and would defeat the purpose of an integrated database incorporating many problems. Table 1 shows the four different solutions (education change; social movements; business strategies and government involvement), posted as a question, and for each solution, four specifics.

Step 4 – Create the Self-profiling Classification Questions to Learn the Respondent's Gender, Age, and Optional Behavior Provided by the Third Question

The third question in the self-profiling question. The actual topic of the study was given (Loss of hope), and then the four alternatives. The same format applied to all studies. Only the topic of the actual study change.

Table 1: The four types of solutions, and the four specific solutions for each type.

	Question A: What education changes need to be instilled?
A1	Embedding the issue in school curriculum
A2	Promote the voice of young students
A3	Recruiting teachers who are activists in their communities
A4	Promote educational messaging with subject matter experts
	Question B: What social movements need to start?
B1	Create self-help movements
B2	Start a protest and improve conditions within the government
B3	Create a riot to overthrow the government
B4	Promote social media activism
	Question C: What business strategies need to be implemented?
C1	Put company executives on the ground floor to understand and act on the issue
C2	Rely on business innovation to provide the solution
C3	Embedding issue within business operations
C4	Big spending philanthropic initiatives by businesses
	Question D: What government involvement needs to happen?
D1	Create laws and legislation to prevent the issue
D2	Provide government funding
D3	Public outreach through mailers and mass messaging
D4	Incentivize behaviors...tax breaks

Preliminary Question: What is the most effective approach to solve the problem of Loss of hope - People who have no hope that anything they do will help their lives

1=Education Changes

2=Social Movement

3=Business Strategies

4=Government Rules

Step 5 - Create the Test Combinations Using Experimental Design

Conventional research works with single ideas (idea screening or promise testing), or with completed 'concept' or even advertisements. Typically, one has no way of knowing what ideas will win, or how a concept will score. The astute researcher limits risk by narrowing down the effort, generating a good sense of what answer will be obtained, and choreographing the research to accept or reject the ingoing hypothesis. Thus, in the end, most research is not so much to 'discover' and to confirm, presumably because most researcher is subtly based upon a 'pass/fail' system.

Mind Genomics is different. Mind Genomics screens ideas, combinations of the elements in Table 1, almost metaphorically in the way an MRI takes pictures of the tissue from different angles, and then combines these pictures at the end to produce an in-depth visual representation. No one picture is correct. Rather, it is the many different combinations which are processed to generate a pattern. The MRI does not 'test,' but rather recreates from different angles. Mind Genomics uses the same approach, albeit metaphorically, by testing different combinations of the elements, getting reactions, and from the patterns of the reactions, showing elements which drive solutions to the problems, and elements which did not.

Mind Genomics works by experimental design, systematic combinations of answers to problem. The problem is presented, and then the Mind Genomics program presents different combinations of these solutions. The respondent simply rates the combinations on a scale. The respondent ends up doing the rating by intuition, rather than trying to guess what the right answer is

Step 6 – Create an Orientation Paragraph, Introducing the Respondent to the Topic

For most research it is not necessary to create a long-set up. A short paragraph, even a single sentence will do the job. For this study we put together a more general paragraph, which could work with the different problems. The orientation ended with the specific problem, here shown in italics, but in normal font in the actual study. Figure 2, left panel, show how the orientation is typed into the BimiLeap template. The actual text follow, with the topic of the study in bold.: *America is full of unsolved issues. You will see a list of possible actions to solve a problem: Loss of hope - People who have no hope that anything they do will help their lives. Please use the scale below to tell us what will happen when the solutions are applied to deal with this problem: Loss of hope - People who have no hope that anything they do will help their lives*
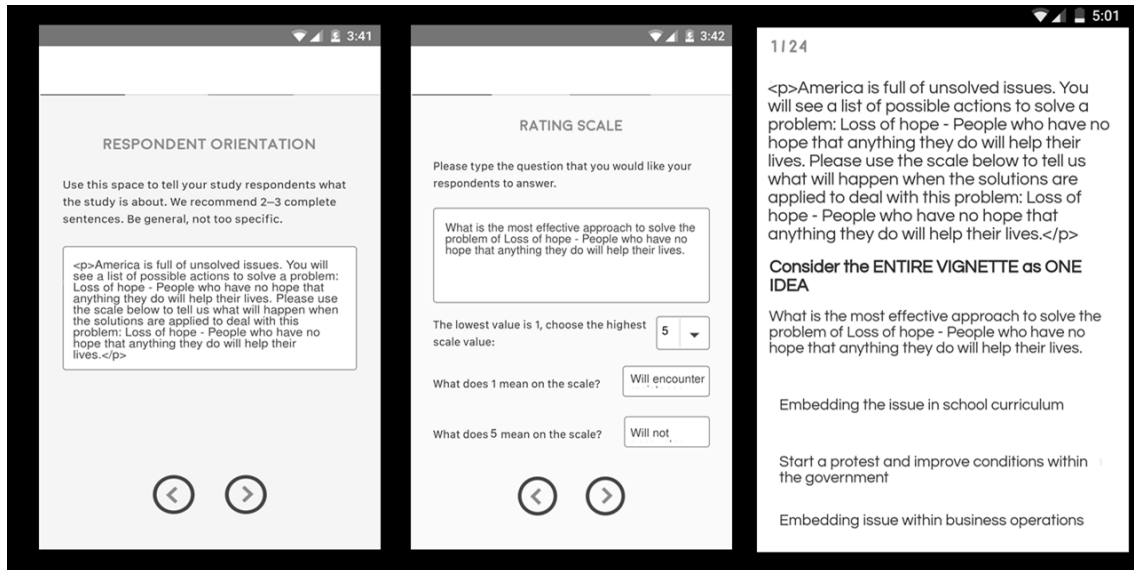
Figure 2: The template acquisition form for the orientation (left panel), the rating scale (middle panel), and an example of one of the vignettes (three elements; right panel).

Step 7 – Create the Rating Scale

The rating scale makes up five labelled points. The enables the researcher to deal with two dimensions, resistance (no/yes), and work (no/yes). Figure 2 (middle panel) shows the templated screen to type in the rating scale.

What is the most effective approach to solve the problem of *Loss of hope - People who have no hope that anything they do will help their lives.*

1=*Will encounter resistance … and… Probably won't work*

2=*Will not encounter resistance… but ... Probably won't work*

3=*Can't honestly decide*

4=*Will encounter resistance… but ... Probably will work*

5=*Will not encounter resistance ... and… Probably will work*

Step 8: Present Each Respondent with 24 Vignettes

Figure 2 (right panel) shows an example of a vignette. Each respondent begin with the self-profiling classification, then read the orientation page, and then rated 24 vignettes on the 5-point scale. The program presented the combination, acquired the rating (5-point scale), and the number of seconds, to the nearest tenth of second, between the time the vignette was presented and the time that the response was made.

Step 9 – Prepare the Data for Statistical Modeling

A key benefit of Mind Genomics is 'design thinking.' Rather than getting data and testing hypotheses, Mind Genomics is set up to create a database. The data itself forms rows of data. Each respondent generates 24 rows of results, with the following columns.

a. Columns 1-3: The columns record the topic, the respondent identification code, the age, gender, and answer to the third classification questionnaire. These are the same for the 24 rows of data for that respondent.

b. Columns 4-19: There are 16 elements that could be incorporated into a vignette. Each of the next 16 columns corresponds to an element, with the value '1' inserted when the element appeared in that vignette, else the value '0' inserted when the element was absent from that vignette. The experimental design prescribed which set of 2-4 elements would appear. Thus, any row would show two, three, or four '1's,' and the rest 0's

c. Columns 20-22: The respondent rated the vignette on the 5-point scale. The next three columns show order of test (1-24) of the vignette, the rating assigned by the respondent (1-5), and the number of seconds to the nearest 10th of second between the appearance of the vignette and the respondent's rating (0-8 seconds; all times > 8 seconds were truncated to 8).This is called the RT, the response time

d. Columns 23-27: The datafile was manually reshaped by augmenting it with five new variables (R1-R5), showing which rating was assigned. For example, when the respondent to rate the vignette '5', R5 took on the value 100 (with a vanishingly small random number added), whereas R4, R3 R2 and R1 each took on the value 0 (also with a vanishingly small random number added). The random number is a prophylactic measure for the downstream regression models, yet to come.

e. Columns 28-31: Four new variables were created, allowing the database to feature a single variable emerging from both instances of an answer. For example, the phrase 'Probably will work' appears in R4 and R5. Thus, R45 takes on the value '100' (plus the vanishingly small random number) when the rating was either 4 or 5, respectively. R45 takes on the value 0 (Plus the vanishingly small random number) when the rating was 1,2 or 3, respectively. The four newly created variables of this type are:

R45 Probably will work

R25 will not encounter resistance

R12 Probably won't work

R14 will encounter resistance

Step 10 – Use Clustering to Create New to the World Mind-sets, Individuals Who View the World the Same within this Specific Framework of Problems and Solutions

At the start of the on-line experiment, the respondent completed a small, three-question self-profiling questionnaire, to record gender, age, and select of which approach would be the best way to solve the problem. A hallmark of the Mind Genomics approach is to let pattern of responses to the granular issue generate possibly new-to-the-world groupings of respondents, not based on who they are, but based on how they respond to the granular issues (here solutions to problems). These groups are mind-sets. The respondent may or may not even be aware of belonging to a mind-set, but the response pattern to the 24 vignettes will reveal that membership, after the responses to the vignettes are deconstructed into the part-worth contributions of each of the 16 elements.

Clustering is a well-accepted group of statistical methods which divide objects into non-overlapping groups based upon patterns of features shared by the objects. In our case the pattern of features will be the degree to which each of the 16 elements drives the response. The elements will be coded as 0's and 1's, in the database, and the criterion variable will be R45, the rating of 'probably will work'. The analysis, purely mathematical, will create a profile of 16 numbers (coefficients) for each respondent, each coefficient attached to one of the16 elements. The clustering program [12] will put respondents into two groups, and then three groups, based strictly on mathematical criteria, not judgment. It will be the job of the researcher to select which set of groupings makes sense (two groups vs. three groups). The criteria will be parsimony (the fewer the number of groups or clusters, the better), and interpretability (the groups must make sense).

The novel approach here is that the clustering will be done on the coefficients of all 257 respondents. Thus, the clustering will look at the way the respondents feel a problem can be solved, with the problem varying by experiment, and clearly stated at the start of the experiment. Psychologists call this process priming [13].

The method for creating clusters follows the rules of statistics. The total data-set includes five studies, slightly more than 50 respondents per study. Recall that the 24 vignettes for each respondent were laid out by an experimental design. Even though the combinations were different for each respondent, the mathematical structure was the same. This is called a permuted design [14]. The benefit of the individual level experimental design it that it allow the researcher to use OLS (ordinary least-squares) to relate the presence/absence of the 16 elements either to the rating, to the binary transformed rating, or to response time

When OLS regression is applied to the data, one respondent at a time, using the option of 'no additive constant,' the individual level regression appears as:

Binary Response (R45)=k_1(A1)+k_2(A2)+... k_{16}(D4)

The foregoing equation expressed how the 16 different answers shown in Table 1 can be combined to estimate the rating of R45, the newly created variable 'probably will work.' Thus, the regression analysis extracts order from the data, allowing patterns to appear. High coefficients suggest that when the element in inserted into the vignette, the rating is likely to be either 4 or 5, both corresponding to probably will work. Low coefficients suggest that when the element is inserted into the vignette, the rating is likely not to be 4 or 5.

The individual level, and for that matter the group models, do not have an additive constant. This revised form allows the direct comparison of the 16 elements. It is vital to be able to compare the elements, side by side, across groups. The additive constant is more correct statistically, but makes the comparisons difficult. Thus, for this r set of analyses we choose not to use the additive constant, even though the model will not fit the data as well.

The five studies were treated identically, considered simply as part of a one big study. To the regression analysis the structure of the inputs and output was identical. At the end of the regression analysis the result was a data matrix forming 16 columns, one for each element, and 257 rows, one for each respondent. The numbers in the data matrix were the coefficients.

A k-means clustering program divided the 257 respondents into two groups and three groups based upon the distances between the respondents. Respondents separated by large distance, to be defined now, were put into different clusters, or mind-sets The 'distance' between people was operationally defined as (1-Pearson Correlation), computed on the 16 coefficients of pairs of respondents, not matter whether they were i the same study or different studies. the structure of their data allowed that.

The Pearson correlation shows the strength of a linear relation between two objects (e.g., respondents). The value of R varies from+1 through 0 to -1. R takes on the highest value, 1, when the two objects are perfectly linearly related. R take on the lowest value, -1, when two objects are perfectly inversely related. The distance between two people goes from a low of 0 when the 16 pairs of coefficients generate a Pearson R of+1 (D=0), to a high of 2 when the 16 pairs of coefficients generate a Pearson R of -1 (D=2).

Step 11 – Create Group Models, Incorporating All the Data from a Group

The final modeling consist of creating a general model which uses all 16 elements as predictors, as well as study, gender, age group, belief in what is the best way to solve the problem, and finally mind-set. The modeling thus puts all the variables on the same footing, allowing the researcher to instantly understand the contribution or driving power of each element or respondent feature to three selected dependent variables. These three variables are R45 (probably will work), R3 (can't make a decision), and RT (response time in seconds).

The general model is expressed as:

Dependent Variable=k_1(A1)+k_2(A2)+...

k_{16}(D4)+k17(College Expenses)+k18(Economic Gap)+k19(Loss of Hope)+k20(Poverty)+k21(Social Security)+k22(Female)+k23(Male)+k24(Age 17-29)+k25(Age 30-49)+k26(Age 50-64)+k27(Age 65-80)+k26(Business)+k27(Education)+k28(Government)+k29(Social Movement)+k30(Mind-Set 1)+k31(Mind-Set 2)+k32(Mind-Set 3)

The foregoing equation is easy to estimate, even for large data sets. It is important to keep in mind that the 16 elements (A1-D4) were designed to be statistically independent and thus always appear in the equation. Not so, however, with the other variables. In every regression model, exactly one of the classifications from each group will be missing, and given a value 0 by the regression. That is because the coefficients for the classification features are relative, not absolute. Thus, when looking at males versus females, there is a variable called male, and another variable called female. One of them will have a coefficient showing its relative contribution (viz. female). The other will be set to 0 (viz. male)

How do the Respondents Distribute Across the Different Classification Criteria?

Table 2 shows the distribution of the respondents across the five studies, each column corresponding to a study. The rows correspond to the distinct groups into which a respondent could be put, either from the up-front self-profiling classification, or from the clustering into three mind-sets.

What is the Pattern of Ratings Assigned by the Respondents in the Separate Groups?

Our first analysis focuses on the pattern of ratings, something that would be the natural first step of any research. We have five ratings (R5, R4, R3, R2, R1, the simple five point scale), as well as four combining scales: Probably will work (R45), Won't encounter resistance (R25),

Probably won't work (R12), and Will encounter resistance (R14), respectively.

Faced with the data, and absent consideration of the underlying experimental design, the standard analytics would begin by compiling a list of frequencies of ratings by key subgroups (Table 3). After doing that, the typical analysis might look for departures, such as groups in the studies seeming to depart from the general pattern.

This surface analysis looks at the pattern for the Total Panel versus the pattern for a specific group, such as the study topic 'Loss of Hope', which seems aberrantly positive. The surface analysis provides observations, but little in the way of deep insight.

How the Ratings Change with Repeated Evaluations

Conventional research often asks a limited number of questions perhaps in a randomized order to forestall order bias. The data from these five studies across 50+respondents and 24 vignettes per respondent allow the researcher to get a sense repeating the same task 24 times. The skeptic would say that it is impossible, and that no one can be consistent across 24 vignettes. That skepticism brings up the question of just what happens when the respondent continues to focus on the same topic for 24 vignettes; the vignettes are all different from each other, so we cannot look at the ratings for the same vignette over time. But we can look at the average ratings of vignettes in the same position of time, to see whether we can find a pattern of average vs. time, recognizing of course that the no two vignettes are alike. The issue is whether there is a noticeable position effect.

Table 3: Percent of responses for each group assigned to original ratings (adds to 100), and then both positives (Probably work, Encounter no resistance), and negative (Probably not work, Encounter resistance)

5-Point Rating	Original Rating					Positive		Negative	
	R5	R4	R3	R2	R1	R45	R25	R12	R14
Total	16	25	27	17	16	41	33	33	41
College Expenses	12	27	27	14	20	39	26	34	47
Economic Gap	13	22	23	18	24	34	31	43	46
Loss of Hope	26	25	23	14	11	51	40	26	37
Poverty	14	28	25	18	15	43	32	33	44
Social Security	16	21	36	18	10	37	33	28	31
Female	16	24	29	17	15	40	33	32	39
Male	16	27	23	16	18	43	32	34	45
Age 17-29	19	28	30	15	8	47	34	23	36
Age 30-49	18	24	26	19	14	42	36	33	38
Age 50-64	14	19	25	17	25	33	31	42	44
Age 65-80	12	28	26	15	19	39	27	34	47
Business	17	23	25	16	19	40	33	35	42
Education	13	24	28	17	18	38	30	35	43
Government	17	22	26	19	16	39	36	35	38
Social Movement	18	29	27	15	12	47	33	26	40
Mind-Set 1	15	24	28	17	15	40	33	32	39
Mind-Set 2	15	25	24	17	19	40	32	36	44
Mind-Set 3	18	25	28	15	13	44	34	29	39

Table 2: Distribution of respondents across the five studies.

	Total	Study topic				
		College Expenses	Economic Gap	Loss of Hope	Poverty	Social Security
Total	257	53	51	51	50	52
Female	178	36	37	32	30	43
Male	79	17	14	19	20	9
Age 17-29	59	8	11	9	16	15
Age 30-49	90	9	13	18	14	36
Age 50-64	49	16	15	11	7	0
Age 65-80	59	20	12	13	13	1
Business	48	8	17	2	13	8
Education	89	28	14	24	12	11
Government	54	10	11	6	8	19
Social Movement	66	7	9	19	17	14
Mind-Set 1	97	23	23	15	18	18
Mind-Set 2	95	16	20	22	20	17
Mind-Set 3	65	14	8	14	12	17

To understand the issue of stability with repeated exposure to the same problem we looked at the average rating by position. Rather than looking at 24 positions, we reduced the 24 positions to six by creating six sets of positions (e.g., 1-4, 5-8, et.) and then averaging the four ratings for each respondent to generate six new 'ratings'.

The foregoing analysis allows us to create averages of ratings for each of the six orders, doing so for all respondents in a study, and by each of the five studies. Figure 3 shows the scatterplot of average rating of the 5-point scale versus the new set of six positions. There is a clear order effect, stronger for some (e.g., College Expenses and Economic Gap, less clear for others such as Loss of Hope, Poverty and Social Security). The reason for the differences of average ratings by order of testing is not clear because the five studies were done in the same way.

The change in average rating is important to deal with. It is not usually addressed in conventional research, where the topic is only broached once, and rated. There is nothing to discuss in Figure 3, because we only have a surface measure. However, we deal with Figure 3 as part of a later analysis.

Creating Enhanced Models for the Study Using OLS Regression

In Step 11 above we presented the expression for the enhanced regression model, considering both the elements, as well as the study, gender, age, selected belief about the best solution, and mind-set. The 16 elements are presented as 0's and 1's, the remaining factors (study through mind-set) as category variables which can be deconstructed into separate variables.

As noted above, the equation is:

$Dependent\ Variable = k_1(A1) + k_2(A2) + ... k_{16}(D4) + k_{17}(College\ Expenses) + k_{18}(Economic\ Gap) + k_{19}(Loss\ of\ Hope) + k_{20}(Poverty) + k_{21}(Social\ Security) + k_{22}(Female) + k23(Male) + k24(Age\ 17\text{-}29) + k25(Age\ 30\text{-}49) + k26(Age\ 50\text{-}64) + k27(Age\ 65\text{-}80) + k26(Business) + k27(Education) + k28(Government) + k29(Social\ Movement) + k30(Mind\text{-}Set\ 1) + k31(Mind\text{-}Set\ 2) + k32(Mind\text{-}Set\ 3)$

We run the regression equation by total in the next analysis. In the appendix, we present the parameters of the model by study, by gender, by age, by belief in the best solution, and by mind-sets, as well as by the first and last set of vignettes (to deal with the issue of just what changes as the person evaluates the vignettes)

Table 4 shows the coefficients of models for R45 (Probably work), R3 (Cannot answer), and RT (Response time), respectively. Our first set of analyses focuses only on the coefficients of the 16 elements.

The first data column, labelled RT45 corresponds to the coefficients for the rating of 'probably can solve,' viz., R4 and R5 combined. Surprisingly, eight of the 16 elements generate coefficients of 12 or higher. There are surprises, such as B2 (create a riot to overthrow the government.) This element might not have appeared had the respondents simply rated what ideas would lead to a possible solution, presumably because of an 'internal editor' which tries to be politically correct, and automatically attach a negative response to the element. It is only because the element is embedded in mixture of other elements that the respondent becomes far less capable to be politically correct, simply because it is impossible to be so when confront with what sees a 'blooming buzzing confusion.' The analogy here might be the emergence of negative qualities when a respondent interprets a Rorschach blot. Negative ideas are not easily suppressed in the narrative.

C3 Embedding issue within business operations

B1 Create self-help movements

C4 Big spending philanthropic initiatives by businesses

D1 Create laws and legislation to prevent the issue

C2 Rely on business innovation to provide the solution

B4 Promote social media activism

B3 Create a riot to overthrow the government

D4 Incentivize behaviors...tax breaks

The second column, labelled R3, shows the strong elements driving 'I cannot decide'.. There are no strong performers, viz elements which generate coefficients of+12 or higher. There are two elements which come close. However these are elements which confuse respondents. We would not have really known that, except for the power of this emergent dataset that we are creating.

B1 Create self-help movements

D3 Public outreach through mailers and mass messaging

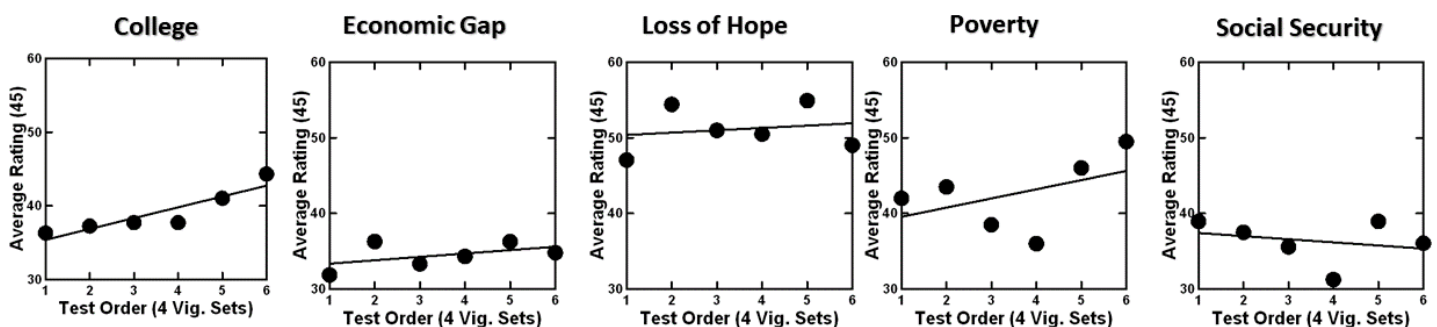The third column, labelled RT, shows the reaction time ascribable to each element. The Mind Genomics algorithm

Figure 3: Change in the average rating over the 24 vignettes by each of the five studies.

Table 4: Models for total panel relating the presence/absence of the 16 elements, the different self-profiling classifications, and the topic study to R45 (probably solve), R3(cannot decide) and RT (response time in seconds).

	Total Panel	R45	R3	RT
C3	Embedding issue within business operations	15	6	1.0
B1	Create self-help movements	14	10	1.2
C4	Big spending philanthropic initiatives by businesses	14	7	0.9
D1	Create laws and legislation to prevent the issue	13	6	0.9
C2	Rely on business innovation to provide the solution	12	6	1.0
B4	Promote social media activism	12	8	1.1
B3	Create a riot to overthrow the government	12	9	1.2
D4	Incentivize behaviors...tax breaks	12	9	0.9
A2	Promote the voice of young students	11	8	0.9
A4	Promote educational messaging with subject matter experts	11	6	1.0
A1	Embedding the issue in school curriculum	11	6	0.9
B2	Start a protest and improve conditions within the government	11	8	1.2
C1	Put company executives on the ground floor to understand and act on the issue	11	7	0.9
A3	Recruiting teachers who are activists in their communities	11	7	0.8
D2	Provide government funding	10	7	0.8
D3	Public outreach through mailers and mass messaging	9	10	0.8
Age	Age 17-29	7	2	-1.2
Age	Age 30-49	3	-4	-0.6
Age	Age 50-64	-8	0	1.0
Age	Age 65+	0	0	0.0
Gender	Female	-1	3	0.3
Gender	Male	0	0	0.0
Pref. Solution	Business	1	-1	-0.1
Pref. Solution	Education	-2	2	-0.1
Pref. Solution	Government	-1	-2	0.3
Pref. Solution	Social Movement	0	0	0.0
Topic	College Expenses	2	-1	0.5
Topic	Economic Gap Disparity	-5	-4	0.7
Topic	Loss of Hope	11	-4	-0.3
Topic	Poverty	0	-2	0.0
Topic	Social Security	0	0	0.0

measured the time from the appearance of the element to the rating, and used that as a dependent measure. Again, these are estimated times needed to read the element and contribute to the decision. The response time can be s a measure of engagement, of reading the information and thinking about it. the response time itself is neither good nor bad, but simply a measure of behavior. The elements which require time to process are those dealing with actions that the person takes:

B1 Create self-help movements

B3 Create a riot to overthrow the government

B2 Start a protest and improve conditions within the government

Our next analysis looks at the contribution of 'the group' of respondents. Following the set of 16 elements (sorted in order) we see four groups. These are the four ways that we divided the respondents, ahead of the study itself. These are age, gender, preferred method of solving problems (all three from the self-profiling classification), and then topic of the study.

One option of each set is always assigned the value 0 because these alternatives in each set are not statistically independent of each other. The respondent must belong to one of the four ages, be one of the two genders, select one of the four preferred methods for solving problems, and take part in one of the five studies. Consequently, incorporating these variables into the regression program meant leaving one of the options out for each group. That option is not estimated in the larger equation, but instead is left out, and in the reporting is automatically assigned the value 0 for its coefficient. It makes no difference which four options are selected. All the coefficients for the options are estimated with respect to the optionally deliberately omitted from the estimation, and automatically assigned the value 0. The four options are Age 65+; Male; Social Movement; Social Security.

The coefficients for each of these four groups can only be compared within the group, not to the other groups, and not to the elements. Nonetheless, we still get a sense of the effects. For example, when it comes to the coefficients for R45, Probably Work, respondents end up generating higher rating when the study topic is "Loss of Hope" with a coefficient of+11. This is independent of all other factors, including elements and ways of classifying the respondent. In contrast 'economic disparity' is the least likely to be solved, at least from these data, with a coefficient of -5.

Looking at the differences between the coefficients for R45, we can conclude that:

1. Age 17-29 is the most positive (+7), whereas age 50-64 is most negative (-8)

2. There are no big differences across the four groups, based on the way they define themselves in terms of what best solves the problem.

3. There is no difference in gender

4. The is a substantial difference in the topic. The coefficient is highest for Loss of Hope (+11) meaning in general people are optimistic that this can be solved. The coefficient is lowest for economic gap disparity (-5) meaning people are least optimistic that this can be solved.

One again it is important to note that this type of information could not be easily obtained from conventional data sources, but becomes a simple byproduct of the data base, trackable over time, and across cultures and events.

We could do the same analysis for R3, the inability to make a judgment. There are no noteworthy group differences in R3, in the way there were for R45.

Finally, the analysis for RT for the age groups suggest that the response for the youngest respondents (age 17-29) is dramatically

faster than the response time for the two older groups (age 50-64, age 65-80). The coefficient for age 17-29 is -0.6. The coefficient for age 50-64 is+1.0 On average, the speed is the response of the older respondents is 1.6 seconds longer for each element.

On the Nature of Micro and Macro Differences among the Three Emergent Mind-sets

The standard analysis by Mind Genomics usually reveals dramatically different, clearly explainable differences across the different mind-sets. Table 5 shows the performance of the elements of these three mind-sets, and the labelled assigned to each. This type of information become increasingly important as the researcher tries to uncover macro pattern among people. It is straightforward to uncover macro patterns when one has commensurate data for all the individuals, as one has here, based on the 16 coefficients.

Traditionally, Mind Genomics stopped after showing the underlying mind-sets and their coefficients. Do we learn any more from knowing the average coefficient in a mind-set, not of the coefficient, but of the different groups?. Are the groups similar, or do the groups differ from each other?

Table 6 shows that there remains heterogeneity across similar groups, even within a mind-set. The variation in coefficients has already been reduced by the clustering, which generated three mind-sets. The remaining variation, that due to the age, gender, preferred solution, and topic, is more of a baseline 'adjustment' value, like the intercept in an equation. One might say that the variables of age, gender, preferred solution of the problem and study topic, respectively are simply additive correction factors of different magnitudes.

The Nature of the Differences between the First and the Last Sets of Four Vignettes

Recall that Figure 3 shows the change in the average rating from the start of the evaluation to the end of the evaluation. Each of the filled circles corresponded to the average of R45 for a set of four vignettes (positions 1-4, 5-8, 9-12, 13-16, 17-20, 21-24). Below figure shows clear that there is an effect. Ordinarily the researcher would report this observation, and move on. The modeling approach allows us to create a full model for each of the six sets of four vignettes. We can create the grand model for the first quartet of vignettes (order 1,2,34), for the last quartet (order 21,22,23, 24) and discover the magnitude of the effect by subtracting the coefficients (Difference=Coefficient for Position 21-24 MINUS coefficient for position 1-4).

Table 7 shows the largest differences for the three dependent variables. There is no need to explain the differences. The intent here is simply to show that these deeper questions can be explore though the database in a way that allows the research to uncover patterns, perhaps unexpected ones, and from that effort generate a working hypothesis.

Discussion and Conclusions

The goal of this paper is to demonstrate a new way of thinking about social issues, one which moves out of the realm of hypothesis testing, and more into the realm of databasing, with objectives to record the citizen's mind in a new way, and as a byproduct lead to

Table 5: Performance of the 16 elements by the three mind-sets.

	Mind-Set 1 of 3 - The solution 'evolves' out of process	R45	R3	RT
A2	Promote the voice of young students	23	6	0.8
A3	Recruiting teachers who are activists in their communities	20	5	0.7
A1	Embedding the issue in school curriculum	20	2	0.8
A4	Promote educational messaging with subject matter experts	19	4	0.8
B3	Create a riot to overthrow the government	17	7	1.1
B1	Create self-help movements	16	7	1.1
C4	Big spending philanthropic initiatives by businesses	16	6	1.0
B2	Start a protest and improve conditions within the government	15	7	1.1
B4	Promote social media activism	15	6	0.9
C2	Rely on business innovation to provide the solution	14	8	1.0
C3	Embedding issue within business operations	13	4	1.0
C1	Put company executives on the ground floor to understand and act on the issue	9	6	1.1
D1	Create laws and legislation to prevent the issue	3	14	0.7
D4	Incentivize behaviors...tax breaks	1	17	0.7
D3	Public outreach through mailers and mass messaging	-3	15	0.6
D2	Provide government funding	-6	15	0.5
	Mind-Set 2 - The solution comes by activism and government action	**R45**	**R3**	**RT**
D2	Provide government funding	26	-2	1.2
D1	Create laws and legislation to prevent the issue	24	-4	1.1
D3	Public outreach through mailers and mass messaging	24	0	1.1
D4	Incentivize behaviors...tax breaks	21	0	1.2
B1	Create self-help movements	18	9	1.4
B2	Start a protest and improve conditions within the government	16	8	1.3
B4	Promote social media activism	15	7	1.2
B3	Create a riot to overthrow the government	11	11	1.4
A4	Promote educational messaging with subject matter experts	7	8	1.1
A3	Recruiting teachers who are activists in their communities	6	8	0.8
A1	Embedding the issue in school curriculum	5	10	1.0
A2	Promote the voice of young students	4	9	1.0
C3	Embedding issue within business operations	4	10	0.9
C1	Put company executives on the ground floor to understand and act on the issue	1	12	0.8
C2	Rely on business innovation to provide the solution	1	10	0.8
C4	Big spending philanthropic initiatives by businesses	1	11	0.8
	Mind-Set 3 - The solution comes from business and government efforts	**R45**	**R3**	**RT**
C3	Embedding issue within business operations	32	3	1.0
C4	Big spending philanthropic initiatives by businesses	29	1	0.9
C2	Rely on business innovation to provide the solution	27	-1	1.1
C1	Put company executives on the ground floor to understand and act on the issue	27	1	0.9
D4	Incentivize behaviors...tax breaks	14	8	0.9
D2	Provide government funding	12	10	0.8
D1	Create laws and legislation to prevent the issue	12	10	0.9
A1	Embedding the issue in school curriculum	10	6	1.0
A4	Promote educational messaging with subject matter experts	8	5	1.0
A2	Promote the voice of young students	8	9	1.1
B1	Create self-help movements	7	14	1.2
D3	Public outreach through mailers and mass messaging	5	14	0.8
A3	Recruiting teachers who are activists in their communities	5	7	1.1
B3	Create a riot to overthrow the government	3	13	1.0
B4	Promote social media activism	2	12	1.1
B2	Start a protest and improve conditions within the government	-3	12	1.1

Table 6: The pattern of coefficients for the total panel, and for the three different mind-sets.

			MS1	MS2	MS3
		Total	**Solution evolves out of the process**	**Solution comes out of activism and government action**	**Solutions comes out of business and government actions**
		R45	**R45**	**R45**	**R45**
Age	Age 17-29	7	5	7	14
Age	Age 30-49	3	6	0	-1
Age	Age 50-64	-8	-6	-8	-11
Age	Age 65+	**0**	**0**	**0**	**0**
Gender	Female	-1	-5	0	4
Gender	Male	0	0	0	0
Pref. Solution	Business	1	-1	3	3
Pref. Solution	Education	-2	1	-2	-8
Pref. Solution	Government	-1	-1	1	1
Pref. Solution	Social Movement	**0**	**0**	**0**	**0**
Topic	College Expenses	2	2	-2	7
Topic	Economic Gap	-5	1	-12	-5
Topic	Loss of Hope	11	11	12	9
Topic	Poverty	0	-3	8	-8
Topic	Social Security	**0**	**0**	**0**	**0**

Table 7: "Large" differences between corresponding coefficients for Positions 21-24 MINUS Position 1-4. The table shows only those major differences, for three dependent variables, R45, R3, and RT.

		Last 4 - First 4
	R45 (Probably work)	
A1	Embedding the issue in school curriculum	14
C1	Put company executives on the ground floor to understand and act on the issue	12
C4	Big spending philanthropic initiatives by businesses	9
Pref Solution	Education	9
C3	Embedding issue within business operations	8
Topic	Poverty	6
A3	Recruiting teachers who are activists in their communities	6
D3	Public outreach through mailers and mass messaging	-6
D2	Provide government funding	-9
B4	Promote social media activism	-11
B3	Create a riot to overthrow the government	-17
	R3 (Can't decide)	
D2	Provide government funding	8
B1	Create self-help movements	-8
B3	Create a riot to overthrow the government	-8
B2	Start a protest and improve conditions within the government	-8
	RT (Response time)	
C1	Put company executives on the ground floor to understand and act on the issue	0.6
B4	Promote social media activism	-0.8
B3	Create a riot to overthrow the government	-1.0
B1	Create self-help movements	-1.1
B2	Start a protest and improve conditions within the government	-0.9

hypothesis generation. The novelty of the approach is the facile, rapid, affordable, and scalable creation of databases having to do with different topics in the same domain.

The It! studies of two decades ago began this effort, but at that time it the value of having the precise elements across all topics was ignored. The It! studies attempted to customize the elements, but at the same time maintain a logical structure spanning all the studies. The result was that each study had to be analyzed separately. The emergence of similar mind-sets across foods [9] was encouraging, but the further analytic power emerging from directly comparability was missing. It was a matter of hoping that the same mind-sets would appear, rather than creating the conditions to use all the data to create a common set of groups spanning all the experiments.

The next logical step can be the expansion of the database across more people within a country, countries beyond the United States, and the creation of the database year after year, or even in an ad hoc way during period of social change. The simplicity and affordability of the database approach as demonstrated here allows for the expansion of this databasing approach to other verticals. In that spirit, the other verticals will feature other topics, and so the topics will change to fit the vertical.

The long term view of the process maybe something like creating a collection of perhaps eight such databases, each dealing with a 'vertical,' viz different facets in of life, each vertical comprising perhaps seven different but precisely parallel studies (topics in the database), each study run with 100 respondents (rather than 50), and study created to be exactly alike and run the same way in 20 countries. This totals 8 (databases/one per vertical) x 7(studies per database) x 100(respondents per study) x 20(countries) or 1,120 studies, each study run with 100 respondents. Verticals could be situations such as conflicts, negotiations, social problems, empowering citizens, enhancing education, and the like. The cost would be minimal (1,120 studies x 400-600 per study as of this writing, Winter, 2022, according to www.BimiLeap.com).

The potential to understand society, its problems, its issues, and opportunities to create a better world through knowledge is the key deliverable from these studies. One might end up with keys which allow groups of people to understand each other, information about communications between hostile parties in conflict situations, along with the ability to update the information, focus that information, or expand the scope as the need arises.

Acknowledgement

References

1. Searles K, Ginn MH, Nickens J (2016) For whom the poll airs: Comparing poll results to television poll coverage. *Public Opinion Quarterly* 80: 943-963.

2. Holwell S (2000) Soft systems methodology: other voices. *Systemic Practice and Action Research* 13: 773-797.

3. Foa UG, Foa EB (1974) Societal structures of the mind. Charles C Thomas

4. PsycINFO Database Record 2016 APA (Foa and Foa, 1974)

5. Moskowitz HR, Gofman A (2007) Selling Blue Elephants: How to Make Great Products that People Want Before They Even Know They Want Them. Pearson Education, 2007.

6. Hunt JD, Abraham JE, Patterson DM (1995) Computer generated conjoint analysis surveys for investigating citizen preferences. In Proceedings of the 4th International Conference on Computers in Urban Planning and Urban Management, Melbourne, Australia, 13-25.

7. Stadelmann S, Dermont C (2020) Citizens' opinions about basic income proposals compared–A conjoint analysis of Finland and Switzerland. *Journal of Social Policy* 49: 383-403.

8. Lilley K, Barker M, Harris N (2015) Exploring the process of global citizen learning and the student mind-set. *Journal of Studies in International Education* 19: 225-245.

9. Moskowitz HR, Beckley J (2006) Large scale concept response databases for food and drink using conjoint analysis, segmentation, and databasing. The Handbook of Food Science, Technology, and Engineering, 2. ed. Y H Hui chapter 59 Taylor and Francis

10. Foley M, Beckley J, Ashman H, Moskowitz HR (2009) The mind-set of teens towards food communications revealed by conjoint measurement and multi-food databases. *Appetite* 52: 554-560. [crossref]

11. Moskowitz HR (2004) Evolving conjoint analysis: From rational features/benefits to an off-the-shelf marketing database. Marketing Research and Modeling: Progress and Prospects (215-230). Springer, Boston, MA.

12. Likas A, Vlassis N, Verbeek JJ (2003) The global k-means clustering algorithm. *Pattern Recognition* 36: 451-461.

13. Molden DC (2014) Understanding priming effects in social psychology: An overview and integration. *Social Cognition* 32: 243-249.

14. Gofman A, Moskowitz H (2010) Isomorphic permuted experimental designs and their application in conjoint analysis. *Journal of Sensory Studies* 25: 27-145.