

## Review Article

# Big Data vs. Big Mind: Who People ARE vs. How People THINK

Howard Moskowitz\*<sup>1</sup>, Attila Gere<sup>2</sup>, Petraq Papajorgji<sup>3</sup>, Daniel Moskowitz<sup>4</sup> and David Moskowitz<sup>5</sup>

<sup>1</sup>Mind Genomics Associates, Inc. White Plains, USA

<sup>2</sup>Szent István University, Faculty of Food Sciences, Department of Postharvest Sciences and Sensory Evaluation, Budapest, Hungary

<sup>3</sup>Professor, European University of Tirana, Tirana, Albania

<sup>4</sup>MindCart AI, Jerusalem, Israel

<sup>5</sup>Independent Researcher, New York, USA

\*Corresponding author: Howard Moskowitz, Mind Genomics Associates, Inc. White Plains, USA

Received: June 05, 2023; Accepted: June 12, 2023; Published: June 19, 2023

## Abstract

This paper presents a new approach to understanding Big Data. Big Data analysis allows to better hypothesize regarding what people think more about certain issues by extracting information on how people move around, what interests them, what the context is, and what do they do. We believe, however, by allowing users also to answer simple questions their interests can be captured more accurately, as the new area of Mind Genomics tries to do. It introduces the emerging science of Mind Genomics as a way to profoundly understand people, not so much by their mind as by the pattern of their reactions to messages. Understanding the way *nature is*, however, does not suffice. It is vital to bring that knowledge into action, to use the information about a person's mind in order to drive behavior, i.e., to put the knowledge into action in a way that can be measured. The paper introduces the Personal Viewpoint Identifier as that tool and shows how the Viewpoint Identifier can be used to evaluate entire databases. The paper closes with the vision of a new web, Big Mind analyzing huge amounts of data, where the networks developed show both surface behavior that can be observed, and deep, profound information about the way each individual *thinks* about a variety of topics. The paper presents a detailed comparison with the Text Mining approach to Big Data in order to understand the advantages of understanding the 'mind' beneath the observed behavior in combination with the observed behavior. The potential ranges from creating personalized advertisements to discovering profound linkages between the aspects of a person and the mind of the person.

## Introduction

When we look at networks, seeking patterns, we infer from the behaviors and the underlying structure what might be going on in the various nodes. We don't actually communicate with the nodes; they're represented geometrically as points of connection. Analytically, we can look at behavior, imposing structural analysis on the network, looking at the different connections—the nodes, the nature of what's being transacted, and the number and type of connections. By doing so, we infer the significance of the node. But, what about that mind in the node? What do we know? What can we know? And more deeply, is that mind invariant, unchangeable, reacting the same way no matter what new configurations of externalities are imposed?

These are indeed tough questions. The scientific method teaches us to recognize patterns, regularities, and from those patterns to infer what might be going on, both at the location and by the object being measured, the object lying within the web of the connections. Mathematics unveils these networks, different patterns, in wonderful new ways, highlighting deeper structures, often revealing hitherto unexpected relations. Those lucky enough to have programs with false colors see the patterns revealed in marvelous reds, yellows, blues, and the other rainbow colors, colors which can become dazzling to the uninitiated, suggesting clarity and insight which are not really the

case. The underlying patterns are clearly not in color, and the universe does not appear to us so comely and well-colored. It is technology which colors and delights us, technology which reveals the secrets.

Now for the deeper question, what lies beyond the network, the edges, inside the nodes, inside the mind lying in the center of a particular connection? Can we ever interrogate a node? Can we ever ask a point on a network to tell us about itself? Does the point remain the same when we shift topics, so the representation is no longer how the nodes interact on one day, but rather interact on another day, or in another situation?

Understanding the environment where a business occurs requires collecting and analyzing massive amount of data related to the potential clients; what they think about the offered products and the level of satisfaction for the offered services/products. The problem of understanding the mind of potential clients is not new; it has been on the focus of marketing researchers for some time to mention a few. One of the most prominent tools to be used for this purpose is *text mining*, defined as a process to extract interesting and significant patterns to explore knowledge from textual data sources. Usually, the collected data are unstructured, i.e., collected from blogs, social media, etc. As the amount of unstructured data collected by companies is constantly increasing, text mining is gaining a lot of relevance [1-7].

Text mining plays a significant role in business intelligence, helping organizations and enterprises to analyze their customers and competitors to make better decisions. It also helps in telecommunication industry, business and commerce applications and customer chain management system.

Usually, text mining combine discovery, text mining techniques, decoding, and natural language processing techniques. The most important elements of this approach are powerful mining techniques, visualization technologies and an interactive analysis environment to analyze massive sets of data so as to discover information of marketing relevance [8,9].

In the world of today, a number of studies suggest that that the efforts to create a technology of text mining have as yet fallen short. Today's (2020) reality in text mining suggest that the effort has performed not as well as was hoped, neither in terms of explicit hopes and predictions, nor the vaster implicit hopes and predictions. Companies which have applied automated analysis of textual feedback or text mining have failed to reach their expectations. emphasize just how hard text mining can be. Research in the area of natural language processing (NLP) encounters a number of difficulties due to the complex nature of human language. Thus, this approach has performed below expectations in terms of depth of analysis of customer experience feedback and accuracy [10].

There are specific areas of disappointment. For example, major obstacles have been encountered in the field of predicting with accuracy the sentiment (positive/negative/neutral) of the customers. Despite what one might read in the literature of consumer researchers and others employing text mining for sentiment analysis, the inability to successfully address these issues has disillusioned some. Some of the disillusionment is to be expected because sentiment analysis must be sensitive to the nuances of many languages. Feelings expressed by words in one language may not naturally translate when the words are translated. Only a few tools are available that support multiple languages. It may be that better feedback might actually be obtained with structure systems, such as surveys.

In this paper we propose a new approach to understand Big Data from the point of view of understanding the mind of the person who is a possible 'node.' We operationally define the world as a series of experiences that might be captured in Big Data, and for each experience create a way of understanding the different viewpoints or mind-sets of the persons undergoing that experience. The effect is to add a deeper level to Big Data, moving beyond the patterns of what are observed, to the mind-sets of the people who undergo the experience. In effect, the approach provides a deeper matrix of information, more two dimensional, the first being the structure of what is being done (traditional Big Data), and the second being the mind-set of the person(s) reacting to this structure. In essence, therefore, a WHAT and the MIND(s) behind the WHAT. We conclude with the prospect of creating that understanding of the MIND by straightforward, affordable experiments, and a tool (Personal Viewpoint Identifier) which allows one to understand the mind of any person in terms of the relevant action being displayed.

## Moving from Analysis of an Object to Interrogating It

We move now from analysis of an object in a network to actually *interrogating* the object in order to understand it from the inside, to get a sense of its internal composition. The notion here is that once we understand the network as externalities *and* understand deep *mind* properties of the nodes in the network, the people, we have qualitatively increased the value of the network by an order of magnitude. We not only know how the points in the network, the people, react, but we know correlates of that reaction, the minds and motivations of these points which are reacting and interacting.

Just how do we do that when we recognize that this mind may have opinions, that the mind may have a desire to be perceived as politically correct, and that, in fact, this mind in the object may not be able to tell us really what's important? How do we work with this mind to find out what's going on inside?

It is at this juncture that we introduce the notion of Mind Genomics, a metaphor for an approach to systematically explore and then quantitatively understand how things are perceived by person(s) using the system. The output of that understanding comprises content (the components of this mind), numbers (a way to measure the components of the mind), and linkages (the assignment of the content and its numbers to specific points, nodes, people in the network) [11,12].

## A Typical Problem – What Should the Financial Analyst Say to Convince a Prospect to Commit?

Lest the foregoing seem to be too abstract, too esoteric, too impractical, let's put a tangible aspect onto the idea. What happens when the point or node corresponds to a person walking in to buy a financial retirement product from a broker whom the person has never met? How does this new broker understand what to say to the person at the initial sales interaction, that first 'moment of truth' when there is a chance for a meaningful purchase to occur? And what happens when the interaction occurs in an environment where the financial consultant or salesperson never even meets the prospective buyer, but rather relies upon a Web site, or a simple outward-bound call-center manned by non-professionals?

The foregoing paragraph lays out the problem. We have our network, nodes connected by the sales activity. By *understanding* the mind of the prospective customer, the financial analyst has a much greater chance of making the sale, in contrast to simply by knowing the age, gender, family situation, income, and previous Web searching behavior—of the prospect, all available from Big Data and grist for the analytic mill. We want to go deeper, into the mind of that prospect.

Psychologists and marketers have long been interested in understanding what drives a person to do something, the former (along with some philosophers) to create a theory of the mind, the latter to create products and services, and sell them. We know that people can articulate what they want, describe to an interviewer the characteristics of a product or service that they would like, often just sketchily, but occasionally in agonizing detail. And all too often this description leads the manufacturer or the service supplier on a wild-

goose-chase, running after features that people really don't want, or features which are so expensive as to make the exercise simply one of wish description rather than preparation for design.

A more practical way runs an experiment presenting the person, this *node* in the system, with different ideas, different descriptions about a product, obtains ratings of the description, and then through statistical modelling, discovers those specific elements in the description which link to a positive response. In other words, run an experiment treating this *node*, this point in a network, as a sentient being, not just as something whose behavior or connections are to be observed as objective, measurable quantities. Looking at the network as an array of connected minds, not connected points, minds with feelings, desires, and opinions, will enrich us dramatically in theory and in practice.

The experiment, or better the paradigm of Mind Genomics, is rather simple. We use a paradigm known as Empathy and Experiment, empathy to identify the 'what,' the content, and experiment to identify the values, the 'important' [13].

Our strategy is simple. We want to add a new dimension to the network by revealing the *mind* of each nodal point. To do so requires empathy, understanding the 'what,' and experiment, quantifying the amount, revealing the structure. Putting the foregoing into operational terms, we will identify a topic area relevant to the node, the person, uncover elements or ideas appropriate to the topic, and then quantify the importance of each element. After Empathy uncovers the raw

materials, the elements, Experiment mixes and matches these elements into different combinations, obtains ratings of the combinations, and then estimates how the individual elements in the combination drive the response.

The foregoing paragraph described an experiment not a questionnaire. Rather, we infer what the person, the node, wants by the pattern of responses and from behavior we determine what elements produce positive responses and what elements produce negative responses [14].

### Putting the Emerging Science of Mind Genomics into Action – Setting Up a Study and Computing Results

The best way to understand the concepts of Mind Genomics, its application to knowledge and to networks, is through an illustration. This paper presents the application of Mind Genomics to create a micro- science about choosing a financial advisor for one's retirement planning. The case history shows the input and practical output of Mind Genomics, how a financial advisor can understand the mind and needs of a customer, identifying the psychological mind-set and relevant points from the very beginning of the interaction. A sense of the process can be obtained from Figure 1. The paper will explicate the various steps, using actual data from a Mind Genomics experiment.

To create and to apply the micro-science we follow the steps below. Although the case history is particularized to selecting a financial advisor, the steps themselves would be followed for most applications.

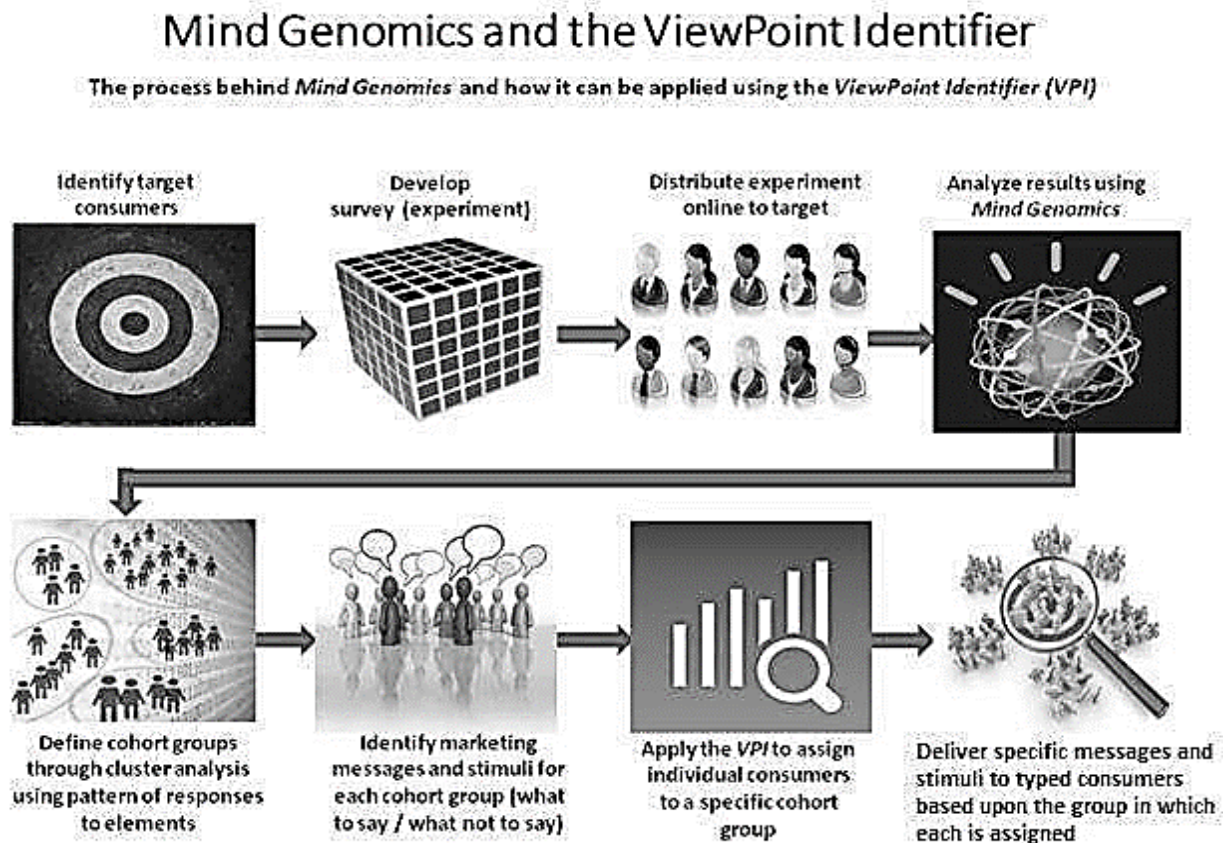


Figure 1: The process of Mind Genomics, from setup to analysis and application. Figure courtesy of Barry Sideroff, Direct Ventures, LLC.

Only the topic area varies.

1. We begin by defining the topic. We also specify the qualifications for the consumer respondents, those who will be part of what might initially look like a Web-based survey, but in reality, will participate in what constitutes a systematic experiment. For our study, the focus is on the interaction of the financial advisor with the consumer, with the specific topic being the sales of retirement instruments such as annuities. The key words here are *focus and granularity*. Specificity makes all the difference, elevating the study from general knowledge to particulars. Granularity means that the data provide results that can be immediately applied in practice.

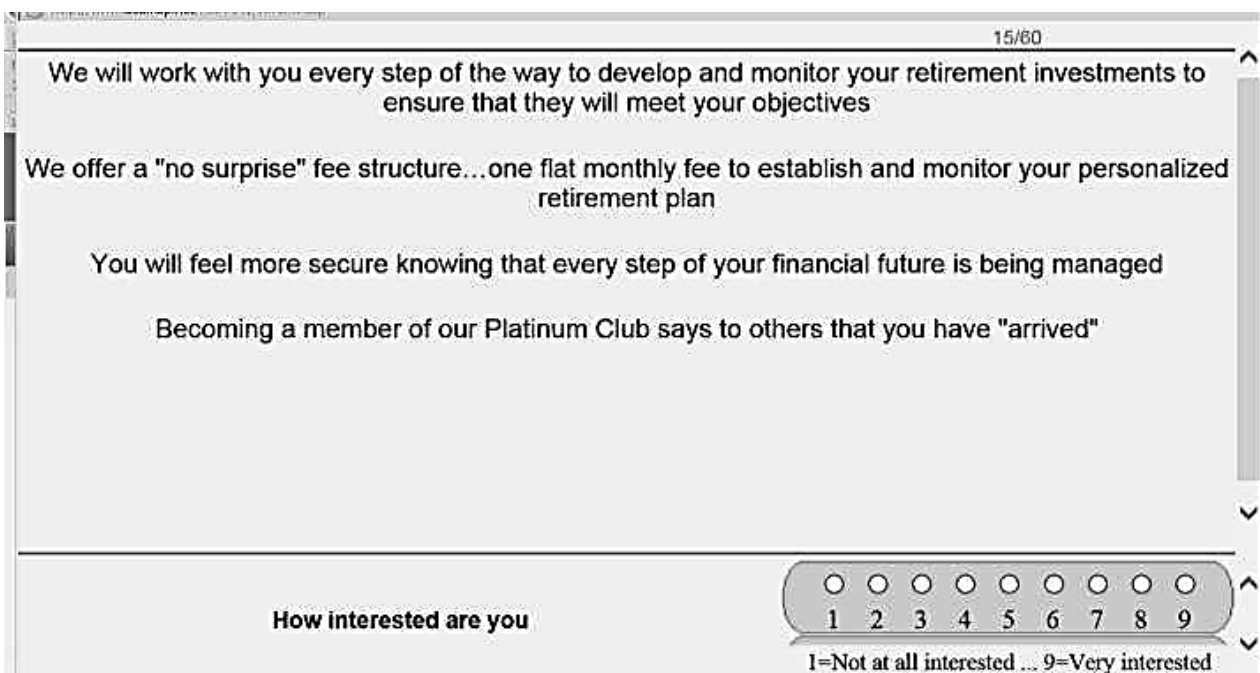
2. Since our focus here is on the inside of the mind, what motivates the person to listen to the introductory *sales message* of the financial planner, we will use simple phrases, statements that a prospective client of the financial analyst is likely to hear from the analyst himself or read in an advertisement. Table 1 presents the set of 36 elements divided into four questions (silos, categories), each question comprising exactly nine answers (elements.) The silos are presented as questions to be answered. This study used a so-called 4x9 matrix (four questions, nine answers per question.) The elements are short, designed to paint a word-picture, and are ‘stand-alone.

**Table 1:** The raw material of Mind Genomics, elements arranged into four silos, each silo comprising nine elements.

	<b>Question A: What will you do, and who will help you?</b>
A1	Wells Fargo...It is never too late to plan for your retirement
A2	Prudential...It is never too late to plan for your retirement
A3	Chase Manhattan...It is never too late to plan for your retirement
A4	Merrill Lynch...It is never too late to plan for your retirement
A5	Tell us when you want to retire, and we will develop an action plan to get you there
A6	Assess your financial health with our retirement planning worksheet...answer some basic questions about your financial present and your goals for the future...then we will recommend some strategies to get you there
A7	We will work with you every step of the way to develop and monitor your retirement investments to ensure that they will meet your objectives
A8	We offer you pointers, recommended reading, and worksheets so you can develop a retirement plan that suits your needs
A9	Your plan will be reviewed annually not just by YOUR consultant but with our entire staff of retirement planning experts...working in collaboration to ensure that your needs are being met
	<b>Question B - How do you make sure that you are in control, and not surprised?</b>
B1	Convenient 24/7 online access to your account and our experts...ask questions...monitor your progress... all online!
B2	We have not forgotten that we are in the relationship business...our retirement planning experts are never more than a phone call away
B3	Our philosophy is based on interpersonal relationships so we will want to meet with you and get to know you...this enables us to develop a better retirement plan for YOU
B4	As you near retirement...you can take steps to improve how long and how well your hard-earned savings will work for you during your retirement years
B5	During your peak earning years, we will help you to manage the changes in your life to the advantage of your longer-term retirement plan
B6	It is never too early to start planning for your financial future...the steps you take today will significantly affect the quality of your life 10, 20 even 40 years from now
B7	We offer a “no surprise” fee structure...one flat monthly fee to establish and monitor your personalized retirement plan
B8	Our fee structure is based on the premise that you only pay for the services you use
B9	We offer a combination fee structure...a low monthly fee plus discounted prices for each transaction
	<b>Question C - How will you feel when you are in control?</b>
C1	With you in charge of your retirement plan...you know if your objectives are being met
C2	Planning for your retirement now means you won't get caught short since you still have time to do something about it
C3	If you don't plan for your retirement...nobody else will...make sure you get to where you want to be
C4	A retirement plan will ensure that you have done all that you can to retire when you want and live the lifestyle you want
C5	Your retirement nest egg may be your single largest asset...invest it wisely with our certified retirement planners
C6	Work with one of our experts to develop your own retirement plan...because Social Security was never meant to be a full retirement program
C7	You will feel more secure knowing that every step of your financial future is being managed
C8	Our Retirement Resource Center has the tools, information, and experts to help you make smart retirement planning decisions
C9	By drafting a retirement plan you will be securing your own financial future
	<b>Question D: What other special 'perks' can you enjoy with us?</b>
D1	You can trust that your local retirement planning expert will be around as you near retirement...they won't be part of some impersonal conglomerate
D2	You can't underestimate the value of your retirement planning expert knowing you, your family, your community
D3	With a nationally recognized retirement planning firm you are never far away from expert advice...no matter where you travel...business or pleasure
D4	A special class of service is available to those who attain Platinum status
D5	Becoming a member of our Platinum Club says to others that you have “arrived”
D6	For a limited time only, meet with one of our retirement planning experts for a free initial consultation
D7	Available exclusively to our VIP customers, a complimentary consultation with someone from the Wills and Trusts department...ask the questions you have been meaning to ask
D8	Planning for your retirement has never been more convenient
D9	Take our Investor Literacy Quiz to identify your areas of strength and weakness...then use our tutorials to fill in your weaker areas



3. A set of 36 elements covers a great deal of ground and typically suffices to teach us a lot about the particular minds of the participants, our respondents, or *nodes* in a web. The particular arrangement of four silos and nine elements is only one popular arrangement of silos and their associated elements. An equally popular arrangement is 6x6, six silos with six elements in each. Recent advances have shown good results with a much smaller set of 16 elements, emerging from four questions, each with four answers (four silos, four elements).
4. Create vignettes, systematically varied vignettes (combinations). The 4x9 design requires 60 different vignettes. Each respondent will evaluate a completely different set of vignettes, enabling Mind Genomics to test a great deal of the possible 'design' space of potential combinations. Rather than testing the same 60 vignettes with many respondents, the strategy of testing different combinations tests more of the possible combinations. The pattern emerges with less error, even though each combination is tested by one, at most two respondents.
5. The combinations, vignettes called profiles or concepts in other published work, comprise 2-4 elements, each element appearing five times. The elements appear against different backgrounds since all the elements vary from one vignette to another. The underlying experimental design, a *recipe book* controls which particular elements appear in each vignette. Although to the untutored eye the 60 different vignettes appear to be simply a random, haphazard collection of elements with no real structure, nothing could be further from the truth. The experimental design is a well-thought-out mathematical structure ensuring that each element appears independently of every other element, repeated the same number of times across each element., This allows us to deconstruct the response to the 60 test vignettes into the individual contribution of each element. Statistical analysis by OLS, ordinary least- squares regression, will immediately reveal which elements are responsible for the rating and which simply go along, not contributing anything.
6. We see an example of a vignette in Figure 2A, program sets up the vignettes remotely on the respondent's computer, presents each vignette, and acquires the rating. The bottom of the vignette shows the rating scale for the vignette. The respondent reads the vignette in its entirety and rates the vignette on the scale. The interview is relatively quick, requiring about 12 minutes for the presentation of the vignettes followed by a short classification questionnaire. The process is standardized, easy, disciplined, and quite productive in terms of well-behaved, tractable data that can be readily interpreted by most people, technical or non-technical alike. As long as the respondent is at least a bit interested and participates, the field execution of the study with respondents is straightforward. The process is automatic from the start of the experiment to the data analysis, making the system scalable. The experiment is designed to create a corpus of knowledge in many different areas, ranging from marketing to food to the law, education, and government. It is worth noting that whereas the 60 vignettes require about 12 minutes to complete, the shorter variation, the 4x4 with 24 vignettes, requires only about 3 minutes.
7. The original rating scale that we see at the bottom of the vignette in Figure 2 is a Likert scale, or category scale, an ordered set of categories representing the psychological range from 1



**Figure 2:** An example of a test vignette. The elements appear in a centered format with no effort to connect the elements, a format which enhances 'information grazing.' The vignette shows the ratings scale at the bottom, and the progress in the experiment at the top right (screen 15 out of 60).

(not at all interested) to 9 (very interested). For our analysis we simplify the results, focusing on two parts of this 9-point scale, with the lower part (ratings 1-6) corresponding to not interested and the upper part (ratings 7-9) corresponding to interested. We re-code ratings of 1-6 to the number 0 and ratings of 7-9 to the number 100. The recoding loses some of the granular information, but the results are more easily interpreted. *Although the 9-point scale provides more granular information, the reality is that managers focus on the yes/no aspect of the results.*

8. The Mind Genomics program also adds a vanishingly small random number to each newly create binary value, in order to ensure that the OLS (ordinary least-squares) regression does not crash in the event that a respondent assigns all vignettes ratings 1-6, or ratings 7-9, respectively. In that case, the transformed binary variables are all 0 or 100, respectively, and the random number adds need variability to prevent a ‘crash.’ The 60 vignettes allow the researcher to create an equation for each respondent Building the model at the level of the individual is a powerful format of control, known to statisticians as the strategy of ‘within- subjects design.’
9. Some of the particulars underlying the modelling are:
  - a. The models are created at the level of the individual respondent, using the well-accepted procedure of OLS, ordinary least squares regression.
  - b. The experimental design ensures that the 36 elements are statistically independent of each other so that the coefficients, the impact values of the elements, have absolute value. The inputs are 0/1, 0 when the element is absent from a vignette, 1 when the elements is present in the vignette.
  - c. OLS uses the 60 sets of elements/ratings, one per vignette, as

the cases. There are 36 independent variables and 60 cases, allowing sufficient degrees of freedom for OLS to emerge with robust estimates

- d. We express the equation or model as: Binary Rating =  $k_0 + k_1(A1) + k_2(A2) \dots k_{36}(D9)$ . For the current iteration of Mind Genomics, we estimate the additive constant  $k_0$ , the baseline. Future plans are to move to the estimation of the coefficients, but ‘force the regression through the origin,’ viz., to assume that the additive constant is 0.
- e. The equation says that the rating is the combination of an additive constant,  $k_0$ , and weights on the elements. The elements appear either as 0 (absent) or as 1 (present), so the weights,  $k_1 - k_{36}$ , show the driving force of the different elements.

### Understanding the Result Through the Additive Constant and the Coefficients

We now look at the strongest performing elements from the equation or model which relates the presence/absence of the elements to the transformed binary rating of 0 (not interested) or 100 (interested). The strongest performing elements appear in Table 2. The table shows all elements which generate an impact value or coefficient 8 or higher for any key subgroup, whether total sample, gender, age, or income, respectively.

1. The total panel comprises 241 respondents. We can break out the total panel in self-defined subgroups, e.g., gender, age, and income. That information is available from the self-profiling classification, a set of questions answered by the respondent after the respondent rated the set of 60 vignettes.
2. The additive constant tells us the conditional probability of a person saying *interested in* what the financial advisor has to

**Table 2:** Strong performing elements for the Total Sample and for key subgroups defined by how the respondent classifies himself or herself. The table presents only those strong-performing elements with average impacts of 8 or higher in at least one self-defined subgroup.

		Total	Gender-Male	Gender-Female	Age - Under 40	Age - Over 40	Income - self-defined High	Income - self-defined high
	Base size	241	44	197	107	134	199	42
	Additive Constant	35	28	36	29	40	34	37
A5	Tell us when you want to retire, and we will develop an action plan to get you there	7	9	6	9	6	7	6
A6	Assess your financial health with our retirement planning worksheet...answer some basic questions about your financial present and your goals for the future...then we will recommend some strategies to get you there	6	6	6	9	4	8	
B1	Convenient 24/7 online access to your account and our experts...ask questions...monitor your progress...all online!	6	8	5	8	5	7	
A7	We will work with you every step of the way to develop and monitor your retirement investments to ensure that they will meet your objectives	5	8	4	6	4	5	3
C2	Planning for your retirement now means you won't get caught short since you still have time to do something about it	3	8	2	2	5	3	4
B6	It is never too early to start planning for your financial future...the steps you take today will significantly affect the quality of your life 10, 20 even 40 years from now	3	8	2	5	2	2	9

say, i.e., assigns a rating of 7-9, when reading a vignette which has no elements (the baseline). *Of course, by design all vignettes comprise elements, so the additive constant is an estimated parameter.* We can use the additive constant as a baseline. For the total panel it is 35, meaning that 35% of the respondents would rate a vignette 7-9. Males are less likely to be positive whereas females are more likely to be positive (additive constants of 28 vs. 36). Those under 40 are far less likely to be positive, those over 40 are more likely to be positive (additive constant of 29 vs. 40). Income makes no difference.

3. Beyond the baseline are the elements, which contribute to the total. We add up to four elements to the baseline to get an estimated total value, i.e., the percent of respondents who say that they would be interested in the vignette about the financial consult were the elements to be part of the advertising.
4. To allow patterns to emerge the tables of coefficient show only those positive coefficients of +2 or higher, drivers of interest. Negative coefficients are not shown.
5. The coefficients for the 36 elements are low. Table 2 shows the strongest elements only, and only elements which generate a coefficient or impact value of +8 for at least one subgroup. We interpret that +8 to mean that when the element is incorporated into the advertising vignette at least 8% more people will rate the vignette 7-9, i.e., say 'I'm interested.' The value +8 has been observed in many other studies to signal that the element is 'important' in terms of co-varying with a relevant behavior. Thus, the value +8 is used here, as an operationally defined value for 'important.'
6. Our first look into the results suggests nothing particularly strong emerges from the total sample. We do see six elements scoring well in at least one subgroup. However, we see no general

pattern. That is, we don't see an element working very well across the different groups. Furthermore, reading the different elements only confuses us. There are no simple patterns.

7. Our first conclusion, therefore, is that the experiment worked at the simple level of discovering what is important, and what is not important. We are able to develop elements, test combinations, deconstruct the combinations, and identify winning elements. The experiment, at least thus far, does not reveal to us deeper information about the *mind(s)* of the respondent. We will find that deeper information when we use clustering in the next section to identify mind-sets.

### Deeper, Possibly More Fundamental Structures of the Mind by Clustering

Up to now we have looked at people as individuals, perhaps falling into convenient groups defined by easy-to-measure variables such as gender, age, income. We could multiply our easy-to-measure variables by asking our respondents lots of questions about themselves, about their attitudes towards financial investors, about their feelings towards risk versus safety, and so forth. Then, we could classify the respondents by the different groups to which they belong, searching for a possible co-variation between group membership and response pattern to elements (Table 3).

The just-described approach typifies the conventional way of thinking about people. We define people as belonging to groups and then search out the linkage between such groups and some defined behavior. Scientists call this strategy the *hypothetico-deductive* method, beginning first with a sense of 'how the world might work,' and then running an experiment to confirm, or just as likely, to falsify that hypothesis. We work from the top down, thinking about what might happen and proceeding merrily to validate or reject that thinking.

**Table 3:** Performance of the strongest elements in the three mind-sets, emerging from the cluster analysis. People in MS1 appear to be the target group to be identified as the promising clients for the financial advisor.

		Total	MS1: Up to date full service retirement planning	MS2: Easy Retirement Planning	MS3: You know Me and Can Help
	<b>Base Size</b>	241	70	71	100
	<b>Additive constant</b>	35	32	40	33
A5	Tell us when you want to retire, and we will develop an action plan to get you there	7	17	8	
B6	It is never too early to start planning for your financial future...the steps you take today will significantly affect the quality of your life 10, 20 even 40 years from now	3	12		2
A9	Your plan will be reviewed annually not just by YOUR consultant but with our entire staff of retirement planning experts...working in collaboration to ensure that your needs are being met	4	11	3	
B1	Convenient 24/7 online access to your account and our experts...ask questions...monitor your progress...all online!		11		
B5	During your peak earning years, we will help you to manage the changes in your life to the advantage of your longer-term retirement plan	6	11	6	3
B7	We offer a "no surprise" fee structure...one flat monthly fee to establish and monitor your personalized retirement plan		11		3

Let's proceed in a different manner, without hypothesizing about how the world works. Let's proceed with the data we have, looking instead for basic groups who show radically different, interpretable patterns. In the world of color this is analogous to looking for the basic colors of the spectrum, red, yellow, blue, which must emerge out of the measured thousands of colors of flowers. Let's work from the bottom up, in a more pointillistic, empirical fashion, emulating Francis Bacon in his *Novum Organum*.

How then do we do this? How do we find naturally occurring groups of people in a specific population who show different patterns of behavior or at least responses for the micro, limited area? That is, we are working with a small corner of reality, one's responses to messages about choosing a financial advisor. It's a limited aspect of reality. How is that reality constituted? Are there different groups of minds out there, groups wanting different features? Are these groups of minds interpretable? To continue with the aforementioned metaphor, can we find the *basic colors* for this aspect of reality, the red/blue/yellow, not of the whole world, but the red/blue/yellow of choosing a financial advisor?

That we have limited our focus to the limited, micro area of messaging for client acquisition by a financial advisor makes our job easier:

1. We are working in a corner, nook, a little region of *reality*. That small region is, however, quite granular. We already have rich material produced by our study. Our study with 36 elements and 241 profiles of impact values tells us how 241 individuals value the individual elements.
2. Focusing only on that small wedge of reality, let us see whether there is a deeper structure, focusing only on the reality of choosing a financial advisor and using only the *mind* of the consumer as a way to organize reality. Continuing our metaphor of colors, we have come upon a new limited aspect of reality.
3. What are the basic dimensions of that new, limited aspect of reality? We have only two ground rules. Parsimony and Interpretability, respectively Ground Rule 1, Parsimony: We should be looking for primaries, the fewer the better, for this new aspect of reality, our *mind of selecting the investment advisor*. Ground Rule 2, Interpretability: We must be able to interpret these primaries in a simple way. They must make sense, must tell a story.
4. The foregoing introduction leads us naturally to our data, our 241 rows (one per respondent), and our 36 columns (one per element). The numbers in the 36 columns are the 36 coefficients from the model relating the presence/absence of the 36 elements to the binary transformed rating. We apply the method of cluster analysis to our 241 rows x 36 columns. We do not incorporate the additive constant into our cluster analysis, because it doesn't give us information about the response to particular elements, the focus of the cluster analysis.

5. Cluster analysis puts our 241 respondents first into two groups, then into three groups, then into four groups, and so forth. These are clusters, which we can call mind-sets or *viewpoints* because they represent different viewpoints that people have about what is important in the interaction with a financial advisor. Furthermore, the word 'viewpoint' emphasizes the psychological nature of the cluster, that we are dealing with the *mind* here, the mind as it organizes one small corner of reality, the interaction with a financial advisor.
6. We end up with a solution suggesting three different viewpoints, as Table 3 shows. These three viewpoints are shown and named by virtue of the strongest performing elements in each viewpoint. The additive constants, our baselines, lie in the small range, and are fairly low in magnitude, 30-40. There is no mindset just ready to spring to attention, willing to buy the services of the financial advisor. That ready-to-act mindset would be identified by a high additive constant.
7. The total sample shows no strong elements. This means that without any knowledge of the mind of the prospect it's unlikely that someone will know what to say, or the right thing to say. Perhaps the strongest message, with a coefficient of +7 (an additional 7% interested in working with the advisor) is the phrase: *Tell us when you want to retire, and we will develop an action plan to get you there.*
8. The real differences come from the elements as responded to by the individuals in the different mind-sets. Our most promising group is Mind-Set 1, comprising 70 of our 241 respondents, or 28%. Use the six strong performing elements and one is likely to win over these respondents.
9. If nothing else but the data in Table 3 are known, how might the salesperson 'know' that she or he is dealing with a prospect from Mind-Set 1, versus knowing that the person is in Mind-Set 2 or Mind-Set 3, the less promising mind-sets, the ones harder to convince? Table 3 simply tells us what to say, precisely, once we find the people, a major advance over knowledge that we began with, but not the whole story. It will be our job to assign a new person with some confidence to one of the three mind-sets, in order to proceed with the sales effort. Hopefully, most of the prospects will belong to Mind-Set 1.

### Finding Viewpoints (Minds) in a Population

The foregoing results suggest that we might have significantly more success focusing on the group of people who are most ready to work with the financial advisor. But how do we find these people in the population? The analysis is data analytics, but exactly what should be done? And, in light of the enormous opportunities available to those who can consistently identify these mind-sets and then act on the knowledge, how can we create an affordable, scalable, 'living' mind-set assignment technology?

We walk around with lots of numbers attached to us. Data scientists can extract information about us from our tracks, whether these



tracks are left by our behavior (e.g., websites that we have visited), by forms that we have filled out and are commercially purchasable (e.g., through Experian or Trans Union or any of the other commercial data providers, by loyalty programs, etc.), or even by questionnaires that respondents complete in the course of their business transactions, medical transactions, and so forth.

All of the available data, properly mined, collated, analyzed, and reported, might well tell us when a person is ready to hire a financial advisor, e.g., upon the occasion of marriage, a child, a promotion, a job change, a move to another city, and so forth. But just what do we say to this particular prospect, the person standing before us in person, or interacting with our website, or even sitting at home destined to be sent a semi-impersonal phone message, email, or letter? In other words, and more directly, *What are the precise words to say to this person?*

Those in sales know that an experienced salesperson can intuit what to say to the prospect. Perhaps the answer is to hire only experienced, competent salespeople, with 20 years of experience. After the first 100 or them are hired, what should be done with the millions of salespeople who need a job, but lack the experience, the intuition, and the track of successes, and who are perhaps new to the workforce? In other words, how do we scale this knowledge of the mind of people, so that everyone can be sent the proper message at the right time, whether by a salesperson or perhaps even by e-commerce methods, by websites instead of salespeople?

The foregoing results in Table 3 show us what to say and to whom, especially to Mind-Set 1.. The problem now becomes one of discovering the mind-set to which a specific person belongs. Unfortunately, people do not come with brass plates across their foreheads telling us the viewpoints to which that person belongs. And there are so many viewpoints to discover for a person, as many sets of viewpoints as there are topic areas for Mind Genomics. The bottom line here is that data scientists working with so-called Big Data might be able to infer that a person is likely to be ready for a financial advisor, but as currently constituted, the same Big Data is unlikely to reveal the mind-set to which the individual person belongs. We have petabytes of data, reams of insights, but not the knowledge, the specificity about the way the *mind* works for any particular, limited, operationally defined topic in the reality of our experience.

We move now to the second phase of our work reported here, discovering the viewpoint to which any person belongs. We have already established the micro-science for the financial planner, the set of phrases to use for each of the three mind-sets uncovered and explicated in a short experiment. We know from our 241 respondents the mind-set to which each person belongs, having established the mind-sets and individual mind-set membership in the group membership by used cluster analysis. How then do we identify any new person, anywhere, as belonging to one of our three mind-sets, and thus know just what to say to that person?

In today's computation-heavy world one might think that the best strategy is to 'mine' the data with an armory of analytic tools, spending hours, days, weeks, months attempting to figure out the relation between who a person is, and what to say, in this small, specific,

virtually micro-world. Once that computation is exhausted, there may be some modest covariation between a formula encompassing all that is known about a person and membership in the mind-set. A simpler way, developed by authors Gere and Moskowitz, called the PVI (personal viewpoint identifier), does the same task in minutes, at the micro-level, with modest computer resources, and with the same granularity as the original Mind Genomics study from which the mind-sets emerged.

In simple terms, the PVI works with the data from the Mind Genomics study, viz., the specific information from which the mind-sets emerged. The PVI system perturbs the data, using a Monte-Carlo system, and over 20,000+ runs, identifies the combinations of elements which best differentiate among the segments. The PVI emerges with six elements, all taken from the original study, and with a two-point rating scale. The pattern of responses to the six questions assigns a new person to one of the three (or two) mind- sets.

Figure 2 shows an example of the introduction to the PVI, which asks for information from the respondent. It will be this information which allows the user of the PVI to create a database of 'minds-sets' of people for future research and marketing efforts. Furthermore, the introduction to the PVI has information about the time when the PVI is being completed (important for future work on best contact times), age, gender, etc. The specific questions can be included or suppressed, depending upon the type of information that will be necessary when the PVI is used (viz., research on the time-of-day dependence of mind-sets, if it actually exists.) As of this writing (2023) the PVI can be accessed at: <https://www.pvi360.com/TypingToolPage.aspx?projectid=213&userid=2018>.

Figure 3 shows the actual PVI portion, comprising three questions about one's current life-stage (what is one thinking about in terms of retirement planning), and then six questions designed to assign the new person to one of the three mind-sets. It is important to realize that instead of requiring weeks and heavy computation, the entire process, from the set-up of the PVI to the deployment, is approximately 20 minutes. Like the work to set up a Mind Genomics experiment, system to create a PVI for that study is 'templated', making it appropriate for 'industrial strength' data acquisition. Several studies can be incorporated into one PVI, with studies randomized, and questions randomized, each study or project requiring only six questions, developed from the elements. The process is automatic and can be deployed immediately with thousands of participants within the hour.

Figure 4 shows the feedback emerging immediately from the PVI. The shaded cell shows the mind-set to which the respondent belongs. The PVI stores the respondent's background information (Figure 2) and mind-set information (Figure 4) in a database. Furthermore, the PVI is set up to send the respondent immediately to a website, or to show the respondent a video relevant to the mind-set to which the respondent has been assigned by the PVI (see Figure 5). Thus, the Mind Genomics system comprising knowledge acquisition by a small, affordable experiment, coupled with the PVI, expands the scope of Mind Genomics so that the knowledge of mind-set membership can be deployed among a far greater population, those who have been assigned to a mind-set by the PVI.

The information I am giving you is about my attitudes towards a specific topic(s). As part of the exercise, I am giving you information about myself. This information will be used to provide better choices in products and services, based on your answers.

I Agree to Participate  I Do Not Agree to Participate

Follow Up for Research and Marketing Purposes \*Required  
 Allow  Not Allow

Day Of Week Taken \*Required  
 MONDAY

Approximate Time Taken \*Required  
 6:00 PM to 9:59 PM

If you were given an Admin code enter it below  
 Enter Admin Code if you were given one

If you were given a Respondent code enter it below  
 Enter Respondent Code if you were given one

Email address \*Required  
 mjihrm@gmail.com

Year of Birth \*Required  
 1944

Country \*Required  
 United States

Phone Number \*Required  
 (914)572-1673

Gender \*Required  
 GENDER

Figure 3: Introductory page to the PVI (personal viewpoint identifier).

**RETIREMENT | PVI 07.20.2020.**

ARE YOU SEARCHING FOR RETIREMENT  
 YES  
 NO

HAVE YOU SPOKEN TO A PROFESSIONAL BEFORE  
 NO  
 DON'T REMEMBER  
 YES

HOW DO YOU LIKE TO BE CONTACTED  
 PERSONAL  
 PHONE  
 TEXT  
 EMAIL

TELL US WHEN YOU WANT TO RETIRE ..WE DEVELOP AN ACTION PLAN  
 SORRY...NOT ME  
 THAT'S ME!

IT IS NEVER TOO EARLY TO START PLANNING FOR YOUR FINANCIAL FUTURE  
 SORRY...NOT ME  
 THAT'S ME!

CONVENIENT 24/7 ONLINE ACCESS TO YOUR ACCOUNT AND OUR EXPERTS...ASK QUESTIONS...MONITOR YOUR PROGRESS...ALL ONLINE!  
 SORRY...NOT ME  
 THAT'S ME!

YOUR PLAN WILL BE REVIEWED ANNUALLY NOT JUST BY YOUR CONSULTANT BUT WITH OUR ENTIRE STAFF OF RETIREMENT PLANNING EXPERTS...WORKING IN COLLABORATION TO ENSURE THAT YOUR NEEDS ARE BEING MET  
 SORRY...NOT ME  
 THAT'S ME!

DURING YOUR PEAK EARNING YEARS WE WILL HELP YOU TO MANAGE THE CHANGES IN YOUR LIFE  
 SORRY...NOT ME  
 THAT'S ME!

"NO SURPRISE" FEE STRUCTURE...ONE FLAT MONTHLY FEE.. ESTABLISH & MONITOR YOUR PERSONALIZED RETIREMENT PLAN  
 SORRY...NOT ME  
 THAT'S ME!

Figure 4: The actual PVI for the study, showing three up-front 'questions' about one's general attitude, and then six questions and a 2-point response scale for each, used to assign the person to one of the three mind-sets.

## Evolving into BIG MIND – The Nature Marriage of PVI-enhanced Mind Genomics with Big Data

Up to now we have been dealing with small groups of individuals whose specific mind-sets or viewpoints in a specific, limited topic area we can discover, and then act upon. But what are we to do when we want to deal with thousands, millions, and even billions of new people? Consider, for example, the points in Figure 6, top panel. Consider these points as individuals. Measurement of behaviors show how these individuals connect with each other at a superficial level, at the phenotypical level. There are many visualization techniques which create the interconnections based upon one or another criterion. And from these visualizations we can ascribe something to the network. We can deduce something about the network and the nodes, although not

much, perhaps. We are like psychologists studying the rat. If only the rat could talk, how much it would say about what it is doing and why? Alas, it is a rat, or perhaps a pigeon, the favorite test subjects of those who follow strict behaviorism, of the type suggested by BF Skinner and his Behaviorist colleagues and student at Harvard University. . (Full disclosure – author Moskowitz was a graduate student in some of Skinner's seminars and colloquia, at Harvard, 1965-1968.)

What happens, however, when we know the mind of each person, or at least the membership in, say, four or ten or perhaps 100 or perhaps 1000 different topic areas relevant to the granular richness of DAILY EXPERIENCE? What deep, profound understanding would emerge if we were to know the network itself, the WHO and BEHAVIOR of

RETIREMENT I	MS1 FULL SERVICE	YOU: UP TO DATE FULL-SERVICE RETIREMENT PLANNING	MS2 MAKE IT EASY	YOU: MAKE THE PROCESS EASY FOR ME	MS3 YOU KNOW ME	YOU: I PUT MY FAITH IN YOU ... PLEASE HELP ME GET THROUGH THE COMPLEXITIES
--------------	------------------	--	------------------	-----------------------------------	-----------------	--

Figure 5: Immediately feedback about mind-set membership.

	QuestionCode	QuestionText	Mindset1	Mindset2	Mindset3
1	xxx0	Retirement			
2	xxx1	MindSet Name	MS1 Full Service	MS2 Easy	MS3 You know me
3	xxx2	MindSet Feedback	You: Up to date full service retirement planning	You: Make the process easy for ME	You: I put my faith in you ... please help me get through the complexities
4	xxx3	Mindset Video	XXX	XXX	XXX
5	xxx4	Mindset Link	XXX	XXX	XXX
6	xxx5	Additive Constant		32	40
7	xxx6	What do YOU think?	That's ME!	Sorry...Not Me	33

Figure 6: Set up template for the PVI, showing the ability to show the respondent a video or send the respondent to a landing page, depending upon the mind-set to which the respondent has been assigned by the PVI.

people, coupled with the structure of their MIND, viz., the ‘MIND OF EACH POINT IN THE NODE!’

Consider Figure 6. The top panel shows the aggregate of people. We know WHO the people are. The bottom panel shows the network, WHAT the people do, how they link to each other. What if now we know WHY for each point, how each point thinks about a set of topics. We create a web of interconnected points and discover some of the commonalities of the points, not based on who the points are or what the points did, but rather how the points think about many relevant topics.

How do we move from Mind Genomics of one topic, say our choice of financial advisor, to many topics in common space, say the space of ‘personal finances’ and then through typing people around the world on a continuing basis, as life progresses and events progress: thousands, not hundreds, and finally millions, tens of millions of people. In essence this ‘project’ creates a true ‘wiki of the mind and society’, empirically sound, extensive, actionable, and archival for decades? In essence, how do we go from a map of nodes to a map of connected minds in the every-day life, and across the span of countries and time? (Figure 7)

To reiterate, our goal is to understand the specific mind-set memberships of each point in the network, where the point corresponds to a person. The big picture is thus millions, perhaps hundreds of millions of points, people, observed two ways, and even expanded a third way to billions of people who have completed the PVI, but who may be ‘imputed’ to belong to a mind-set through look-alikes. This is the DVI, the Digital Viewpoint identifier, explicated in step 3 below:

### 1. Granular Mental Information about Each Node

The minds or at least the pattern of mind-set membership of many people determined through Mind Genomics and the PVI, for a

set of different topic areas. There may be as few as one topic area, or several dozen or even 100 or more topics. This information can be obtained through small-scale Mind-Genomics studies, executed and analyzed within 1-2 hours ([www.BimiLeap.com](http://www.BimiLeap.com)), and followed by an easy-to-deploy PVI ([www.PVI360.com](http://www.PVI360.com)).

### 2. Correlate Behavior Observed Externally with the Underlying Mind-sets

The interactions of nodes with each other, as measured objectively, either by who they are or by how they behave, such as what they view on the Web, what they order, with whom they interact in conversations. This information is readily available today from various sources, known collectively as Big Data.

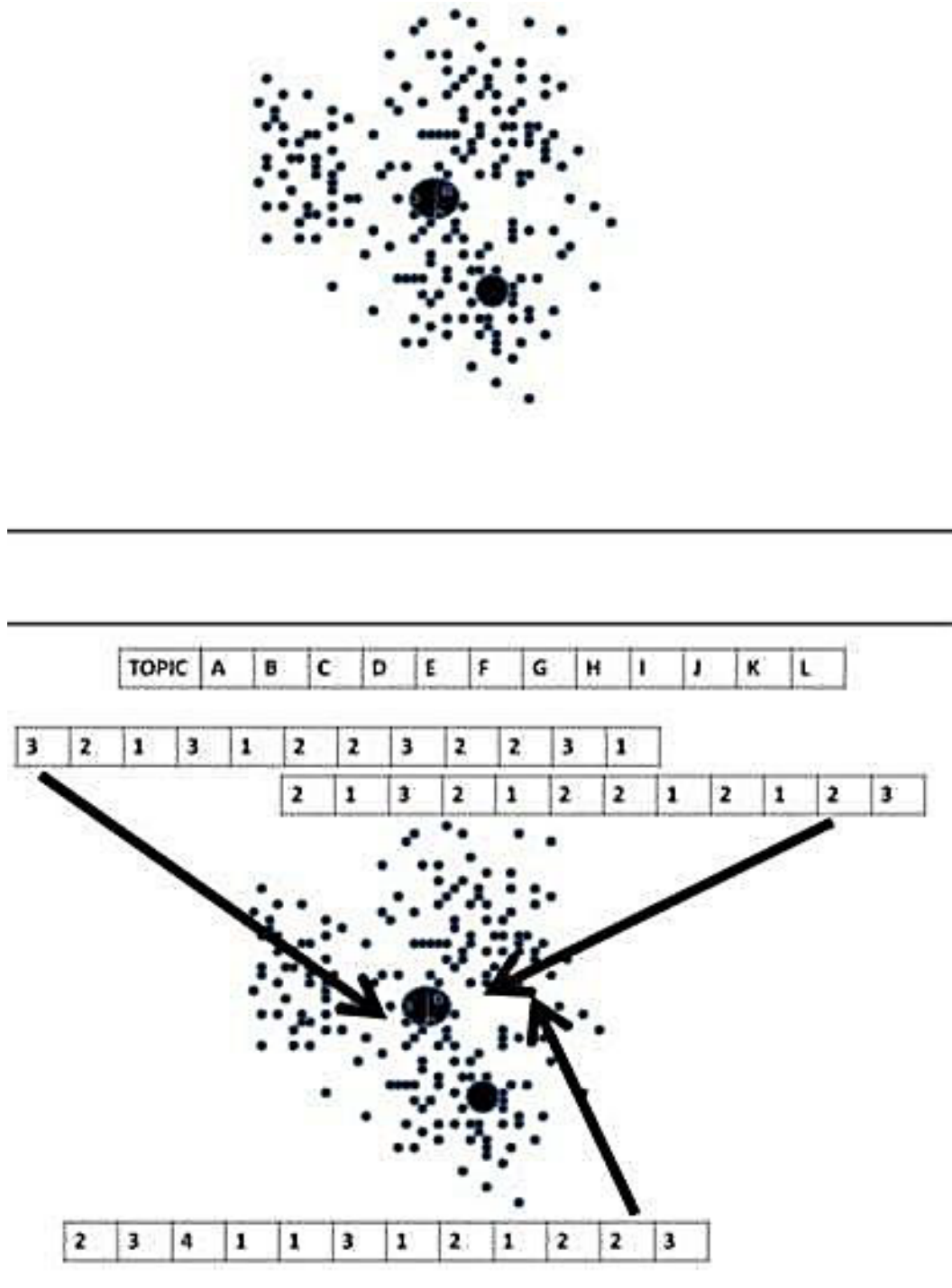
### 3. Expand the PVI (Personal Viewpoint Identifier)

The goal here is to work with 1000 respondents, each of whom provides 5 minutes of her or his time to complete a set of PVI’s on a topic. Let’s choose a number of PVI, say 12. Each PVI of six questions takes about 15 seconds to compete. In three minutes, a person can do 12 PVI’s, comprising 72 questions.

### 4. Augment the Data

Let’s purchase publicly available information about these 1000 known respondents. The goal now is to predict the viewpoints of the 1,000 people on the 12 topics from purchasable data about those 1,000 people. *Once that is done, one has developed a simple predictive model which uses purchasable data to estimate the mind-set membership of a person in each of 12 topic areas from purchasable information that can be readily obtained.* This simple predictive model is the aforementioned DPI, Digital Personal Identifier. It has now become straightforward to create a ‘scoring system’ which moves systematically through the data already available, and ‘scores’ each respondent on 12, 120, or even 1200 different granular topics, to create a true Wiki of the Mind and Society.





**Figure 7:** Example of nodes (i.e., people), perhaps connected by a network. The top panel shows the network of people as points. The bottom panel shows the potential of knowing the mind of each person, i.e., each point in the network.

5. Fast time frame, low cost: Let's consider a simple scenario, the creation of this mass of data for the financial trajectory of a person, from early adulthood to late adulthood, through all the relevant financial aspects. Let's assume 300 different identifiable activities involved in decision-making. The foregoing steps mean that within a period of six months to one year, and some concerted effort, it will become possible, and indeed quite straightforward, to move from say 300 topic

studies to 300 micro sciences and viewpoints, to the creation of 300 digital viewpoint identifiers, to the application of those identifiers, i.e., scoring systems, to the purchasable data of 1-2 billion people. Within the Big Data the data scientist and entrepreneur will have an associated Big Mind, a vector of perhaps 300 numbers underneath each node, each person, each node corresponding to one of those 300 activities. The analytic possibilities emerging from knowing both the



behavior and the mind-set of the behaving organism on 300 (or more) topics can only be surmised. One would not be far off to think that the possibilities are enormous for new understanding of behavior, a possibly new engineering of society.

## Acknowledgments

Attila Gere thanks the support of the Premium Postdoctoral Researcher Program of the Hungarian Academy of Sciences

## References

1. Ordenes FV, Theodoulidis B, Burton, J, Gruber, T, Zaki M (2014) Analyzing Customer Experience Feedback Using Text Mining: A Linguistics-Based Approach. *Journal of Service Research* 17: 278-295.
2. Aciar S (2010) Mining Context Information from Consumer's Reviews. 2nd Workshop on Context-Aware Recommender Systems (CARS-2010)
3. Bucur C (2015) Using Opinion Mining Techniques in Tourism. *Procedia Economics and Finance* 23(Supplement C): 1666-1673.
4. Ritbumroong T (2015) Analyzing Consumer Behavior Using Online Analytical Mining. In *Marketing and Consumer Behavior: Concepts, Methodologies, Tools and Applications* (1st ed, pp. 894-910) IGI Global.
5. Roll I, Baker RS, Alevon V, McLaren BM, Koedinger KR (2005) An analysis of differences between preferences in real and supposed contexts. *User Modeling-Springer Berlin Heidelberg* 367-376.
6. Talib R, Hanif MK, Ayesha S, Fatima F (2016) Text Mining: Techniques, Applications and Issues. *International Journal of Advanced Computer Science and Applications* 7: 414-418.
7. Zhong N, Li Y, Wu ST (2012) Effective Pattern Discovery for Text Mining. *IEEE Transactions on Knowledge and Data Engineering* 24: 30-44.
8. Auinger A, Fischer M (2008) Mining consumers' opinions on the web. In *FH Science Day*. 410;419.Linz, Österreich
9. Fan W, Wallace L, Rich S, Zhang Z (2010) Tapping the power of text mining, *Communications of the ACM*, vol. 49, no. 9, pp. 76-82, 2006. Fatima, F, Islam, W, Zafar, F, Ayesha, S. Impact and usage of internet in education in Pakistan. *European Journal of Scientific Research* 47: 256-264.
10. Fenn, J, LeHong H (2012) Hype Cycle for Emerging Technologies. Gartner.
11. Gofman, A, Moskowitz H (2010) Isomorphic permuted experimental designs and their application in conjoint analysis. *Journal of Sensory Studies* 25: 127-145.
12. Jung, C. H (1976) *Psychological Types: A Revision* (The Collected Works of C.G. Jung) Princeton, NJ: Princeton University Press
13. Moskowitz HR, Batalvi B, Lieberman E (2012) Empathy and Experiment: Applying Consumer Science to Whole Grains as Foods. In *Whole Grains Summit* (2012) (pp. 1-7) Minneapolis, MN: AACCC International.
14. Gofman A (2012) Putting RDE on the R&D Map: A Survey of Approaches to Consumer-Driven New Product Development. In A. Gofman, H. R. Moskowitz (Eds.), *Rule Developing Experimentation: A Systematic Approach to Understand & Engineer the Consumer Mind* (pp. 72-89) Bentham Books.

## Citation:

Moskowitz H, Gere A, Papajorgji P, Moskowitz D, Moskowitz D (2023) Big Data vs. Big Mind: Who People ARE vs. How People THINK. *Internal Med Res Open J* Volume 8(1): 1-13.