A data-driven approach to create a novel, comprehensive personality model using machine learning: the development of the Interpersonality model

Mark Fourman
Chief Data Scientist at Interpersonality Pte. Ltd.
M.A. Mathematics, Oxford University, UK
M.Sc. Microelectronics Systems Design, Brunel University London, UK
Email address: mark@interpersonality.com

ABSTRACT

Our research objective was to build a novel personality type model that is comprehensive in the range of human psychology it covers, simple to understand, and statistically validated for accuracy. We surveyed 144,164 individuals, who completed standard Big Five tests (input variables). We also collected respondents' responses to 9,394 descriptions, covering a wide range of personality and behavioral topics (output variables). In total, we collected 16,181,276 ratings of output variables. Using these data, we applied various machine learning approaches to develop a novel personality type model. We discovered that historical concepts of personality type cannot be used to develop a comprehensive and statistically sound personality type model. In order to be comprehensive and have enough detail to be insightful, the number of personality types needed was much too large. Instead, in our model resulting from this research, we identified 14 separate personality themes (e.g. Extraversion, Conflict, Emotions) with between 2 and 8 styles within each theme. Each person's personality is made up of their style in each of the 14 themes. This allows for great detail and insight while still being simple enough to understand. In total, there are 29,859,840 potential combinations of styles. The ability to generate about 30 million personality reports matches the complexity of human personality better than existing models, which typically only generate reports covering 16 personality types.

Keywords: *Personality, Big Five, IPIP, Interpersonality model*

CONTENTS

1. INTRODUCTION

When we set out on the research project, our goal was to create a more accurate, more comprehensive and more useful personality model that would provide new insights into key areas of human psychology and behavior, specifically: romantic relationships, career path, lifestyle, working relationships and life outcomes.

1.1. Research objectives

We aimed to overcome some fundamental limitations of existing personality models:

- **Dichotomous models** that put people into one of two extremes, without recognizing the fact that most people fall on the midline in most measurable personality traits.
- **Limited scope**, where personality type models only address a small fraction of what makes up a person's personality, and ignore all the rest.
- Creating **false distinctions** that claim to differentiate people where no statistically meaningful differences actually exist.

1.2. Criteria

At the start of our project, we set the criteria for the personality type model we aimed to develop:

- Avoid using the Barnum Effect [1] to write personality profiles that most people would agree with, regardless of their personality type, and instead develop a model where people would strongly agree with what was written in their own profile and disagree with what was written in someone else's profile.
- Provide accurate and useful insight into important areas of personality psychology and behavior, specifically in the areas of romantic relationships, career path, lifestyle, working relationships and life outcomes.
- Comprehensive - not just covering a very few factors of personality.
- Simple enough to be understandable.

2. LIMITATIONS OF EXISTING PERSONALITY MODELS

In this section, we will discuss the limitations of existing personality type models. Before we do so, we would like to provide context for this critique.

Our goal here is not to diminish the contributions that previous work has made to the field of psychometrics, but rather to describe the limitations of previous models in order to identify where there are opportunities for new research and improvement on previous models.

- We especially value the work of Myers and Briggs to develop what is now the Myers-Briggs Type Indicator® [2]. This was one of the first ever personality tests and created a foundation for a century of research and development in psychometrics.
- William Moulton Marston to develop what is now the DiSC® Personality Test [3][4].
- Costa & McCrae for developing the five factor model and the The NEO Personality Inventory, which has evolved into various versions of the Big Five model and tests [5].

Without these previous contributions to the field, our current research would not be possible.

In our critique we often reference the Myers-Briggs® model, because it is the most widely used personality type system and has been the subject of extensive research.

2.1. Dichotomous Models vs. Continuous Models

Almost all personality type models use four personality traits to construct their personality type models, so in this section, we'll assume the personality type model has four traits. (The traits used in different models to construct their personality types differ widely, but almost all models use four traits.)

A few personality models, such as DiSC, provide trait score results on a continuous scale. (Although with the caveat that DiSC uses an ipsative scale, which we discuss below.)

However, most personality models form their type system by dividing people into one or other extreme of a trait - above or below some cutoff point of the trait score in that dimension. For example, MBTI® classifies people into Extraverted or Introverted. They then define the different personality types as combinations of one or other of those extremes in each trait dimension.

With four measured traits in the model, this creates $2^4 = 16$ personality types.

The problem with this approach is that for any measurable personality trait, people fall on roughly normal distributions, with most people being close to the mean, and a few outliers being near the high or low extremes.

- For example, see Bess and Harvey [6] for an analysis of the distribution of people across the four MBTI trait scores.
- Similarly, our own data show roughly normal distributions of scores across the Big Five traits and across all 30 sub-facets of the Big Five.

Thus, defining an arbitrary cutoff point between 'high' and 'low' on a trait dimension creates a false dichotomy between types of people. For example, categorizing people into extraverted *or* introverted and misses the midline 'ambiverts' who are neither highly extraverted nor highly introverted.

As most people are actually midline on most personality traits, this is a serious shortcoming. It means that most existing models fail to describe most people accurately on most of the trait dimensions the models use.

Thus, a key objective in our attempt to develop a new personality model was to ensure that the model accounted for the roughly normal distribution of people across trait dimensions, and accurately describe people who fall near the middle of trait dimensions.

2.2. Limited Scope

The breadth of human psychology and behavior is vast and the number of personality traits that *could* be measured and used to build a personality type model is equally vast.

Obviously, the more traits that are taken into account when building a personality type model, the more opportunity there is to build a *comprehensive* model.

However, there are challenges that come with increasing the number of traits used to build a model. There are two ways of thinking about these challenges, one is historical and one is statistical.

2.2.1. Historical Perspective

Research that led to the development of the Myers-Briggs Type Indicator started in 1917, leading to the MBTI Type Indicator being published in 1962. Marston published his book on the "Emotions of Normal People" in 1928, which is the foundation for the DiSC® Personality System. This research was all conducted long before computing power was generally available. More modern personality type models were also developed before the creation of "Data Science" as a discipline - a term that was first used in 2001.

Thus personality type models had to be restricted to:

- A process for tallying test results to assign a person to their personality type that was simple enough to easily calculate by hand on paper.
- A personality model that was simple enough to be created by a human mind, without recourse to computationally-intensive machine learning.

2.2.2. Statistical Perspective

The limitations of the human brain aside, the complexity of building a personality type model grows exponentially with the number of traits considered.

For example, as existing models create personality types based on combinations of dichotomies, if there are n traits in the model, there are 2^n possible combinations of the two extremes, and thus 2^n personality types. For example, most models use 4 traits and thus have $2^4 = 16$ personality types.

Taking the same approach:

- A personality type system built on the Big Five traits would have $2^5 = 32$ personality types.
- A personality type system built on the six HEXACO traits would have $2^6 = 64$ personality types.
- A personality type system based on the Big Five's 30 sub-facets would have 2^{30} = 1,073,741,824, or just over one billion personality types.
- A personality type system based on the Big Five's 30 sub-facets plus age and gender would have 2^{32} = 4,294,967,296, or just over four billion personality types.

Existing personality type models avoid this complexity by limiting the scope of the traits they measure and report on.

That was a necessary compromise in the historical eras these models were developed.

2.3. False Distinctions

Some personality tests use Likert scales for their test questions, while others use ipsative or forced-choice questions.

Likert scales allow the user to respond on a scale of agreement or disagreement to a statement. The rating scales typically allow responses on an integer scale, for example from -3 (strongly disagree) to +3 (strongly agree).

Ipsative, or forced-choice, questions do not allow for gradations in agreement or disagreement, and prevent neutral responses. Typically these tests either force a choice between two options, e.g. "Which of these two options describes you best?" or forced ranking, e.g. "Rank the following statements in order of how well they describe you."

Research shows that Likert-scales can be used to develop personality tests which make comparisons between different groups, or types of people, while ipsative questionnaires cannot.

To quote Wikepedia [7] :

> While mean scores from Likert-type scales can be compared across individuals, scores from an ipsative measure cannot. To explain, if an individual was equally Extraverted and Conscientious and was assessed on a Likert-type scale, each trait would be evaluated singularly, i.e. a respondent would see the item "I enjoy parties." and agree or disagree with it to whatever degree reflected his/her preferences.
>
> If the same traits were evaluated on an ipsative measure, respondents would be forced to choose between the two, i.e. a respondent would see the item "Which of these do you agree with more strongly? a) I like parties. b) I keep my work space neat and tidy." Ipsative measures may be more useful for evaluating traits within an individual, whereas Likert-type scales are more useful for evaluating traits across individuals.

As we were interested in developing a personality type model that described differences between people we adopted a Likert rating scale of -3 to +3 for our 120 test questions. We cover the choice of our test questions in detail in section 3.4.

In our data gathering (see section 4.2), we asked people to respond to 120 personality test questions and then asked them to rate their agreement or disagreement to additional, optional personality descriptions on the same Likert scale of -3 to +3. The sources of those additional descriptions are described in section 4.2.1.

For example, we tested a wide range of response questions relating to *how* people like to be shown affection. As we used a Likert scale, we were also able to capture and evaluate not only *how* people liked to be shown affection but also *how important* being shown affection was to them.

Dr. Gary Chapman, author of the book The Five Love Languages, explored the same topic, using an ipsative questionnaire [8].

After taking Chapman's Love Languages test, people are told that the results give them accurate information on how they most like to be shown affection.

Technically, that is true. But it misses the forest for the trees.

Using a Likert questionnaire, we discovered that the biggest distinction between groups of people is that some people really do care about receiving demonstrations of affection, while for others it's just not that important. Among those who do place importance on receiving demonstrations of affection, there is one sub-group that places more importance on physical touch than others, but this distinction was much less marked than the differences between people who do and don't care about receiving demonstrations of affection.

We found similar patterns when we investigated other themes such as communication styles, leadership styles and negotiation styles. When we compared our results to other tests that covered similar topics using ipsative questionnaires, we generally found that the biggest distinction was whether people cared about the topic or not, and the distinctions drawn from the ipsative tests were much smaller than their personality test results suggest. Some people are great communicators and others are not, but the minusca of their differences in style were less important. Some people like to be leaders and others prefer to follow, but the differences in leadership styles were small compared to their overall preference to lead or follow. Some people have great negotiation skills and others do not, but the differences in the negotiation styles of good negotiators were much less pronounced.

In general, we conclude that forced ranking tests often produce false distinctions that do not reflect people's real preferences. These tests can indeed be useful and interesting. However, when considering the macro picture of a person's personality, they do not represent significant differences between people.

Having settled on using a Likert scale for our personality test, the next question was, which personality traits would we measure to build our model? And which personality test would we use to assess those traits?

3. THE BIG FIVE TEST

The Big Five is the one of the very few statistically-validated personality trait tests for normal psychology [9]. ("Normal" here as distinct from tests of psychopathology, such as the Minnesota Multiphasic Personality Inventory [10][11], which is a statistically validated and well researched test, but does not measure normal psychology).

The Big Five measures five core traits, which are relatively enduring characteristics that influence behavior across many situations. The traits are Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

The original Big Five test was the NEO Personality Inventory [12]. Since then, several other tests, measuring the same Big Five traits, have been developed. See section 3.4 for more details.

The Big Five breaks each of these five major traits down into six facets [13] within each major trait, creating a total of 30 traits.

3.1. Orthogonal Traits

One major technical advantage of the Big Five is that the five traits have been shown to be orthogonal, i.e. each trait is independent of each other - there are near-zero cross correlations between the different traits when sampled across different people. From a mathematical perspective, this is a huge advantage, as working with an orthogonal input vector space makes it much easier to analyze relationships between inputs (trait scores) and outputs (users' Likert-scale responses to descriptions about their personality).

This is in contrast to many other personality trait models. For example, with the Myers-Briggs test, research has shown that the four traits are not orthogonal, and one of the traits does not even measure opposites, but rather two different things (judging and perceiving) where people can in reality score high for *both* judging and perceiving but the test does not allow this result. [14]

From a data science perspective, using traits that are not orthogonal creates difficult technical problems. Having traits where the two poles are not in fact opposite makes meaningful data analysis practically impossible.

3.2. Test-Retest Reliability

An important factor in choosing a personality test and trait model is the test-retest reliability. A test is said to have high test-retest reliability if there are strong correlations between the test results when a person takes and retakes the same test at different times. Some existing personality tests have been shown to have lower test-retest reliability. For example, with the official Myers-

Briggs test, when people retake the test, two or more of their assigned type preferences change between 10% to 25% of the time [15].

The NEO PI-R and other versions of the Big Five test have been shown to have high test-retest reliability.

> The test-retest reliability of the NEO PI-R has also been found to be satisfactory. The test-retest reliability of an early version of the NEO after 3 months was: N = .87, E = .91, O = .86. The test-retest reliability for over 6 years, as reported in the NEO PI-R manual, was the following: N = .83, E = .82, O = .83, A = .63, C = .79. Costa and McCrae pointed out that these findings not only demonstrate good reliability of the domain scores, but also their stability (among individuals over the age of 30).
>
> Scores measured six years apart varied only marginally more than scores measured a few months apart. [16]

3.3. Cross-Cultural Validity

The final important factor in choosing a personality test and trait model is cross cultural validity. It's important that the test and results are not different from one culture to another.

> Evidence of the NEO scales' stability in different countries and cultures can be considered evidence of its validity. A great deal of cross-cultural research has been carried out on the Five-Factor Model of Personality. Much of the research has relied on the NEO PI-R and the shorter NEO-FFI. McCrae and Allik [17] edited a book consisting of papers bearing on cross-cultural research on the FFM. Research from China, Estonia, Finland, the Philippines, France, German-speaking countries, India, Portugal, Russia, South Korean, Turkey, Vietnam, and Zimbabwe have shown the FFM to be robust across cultures.
>
> Rolland [18], on the basis of the data from a number of countries, asserted that the neuroticism, openness, and conscientiousness dimensions are cross-culturally valid. Rolland further advanced the view that the extraversion and agreeableness dimensions are more sensitive to cultural context. Age differences in the five-factors of personality across the adult life span are parallel in samples from Germany, Italy, Portugal, Croatia, and South Korea. Data examined from many different countries have shown that the age and gender differences in those countries resembled differences found in U.S. samples. An intercultural factor analysis yielded a close approximation to the five-factor model. McCrae *et al.* [19] further reported data from 51 cultures. Their study found a cross-cultural equivalency between NEO PI-R five factors and facets.

3.4. Choice of Big Five Test Questionnaire

There are various versions of the Big Five personality test available in the public domain.

In choosing a test, there are two competing factors: the more questions the test has, the more accurately it can measure a person's trait scores, but the longer the test, the more likely people are to get bored and not complete the test.

The 'long-form' Big Five test measures all 30 sub-facets with 300 questions. We considered this to be too long for a personality test to be used by the general public. Fortunately, John Johnson, developed a shorter, 120-question test whose results compare favorably to the full 300-question test. [20]

Johnson's test is in the public domain and available on the International Personality Item Pool site[1].

3.5. A Trait Model, Not a Personality Type Model

To date the Big Five is just a *trait model*, which provides trait scores as a result, but there is limited research into developing a *personality model* built on the Big Five.

To our knowledge, the only major effort in this area has been by Gerlach *et al* [21]. They identified four personality types of people by finding dense clusters of people in the 5-dimensional trait space. With this, they identified four personality types, which they labeled Role Model, Self-Centered, Reserved and Average. While useful and statistically valid, the four personality type description was not particularly descriptive or comprehensive.

Thus we would describe the Big Five as an accurate and personality *trait* model, but not a personality *type* model, because it does not describe how different trait scores interact with each other to form descriptive personality types.

[1] [1] https://ipip.ori.org/30FacetNEO-PI-RItems.htm

4. BUILDING A NEW MODEL

We took an empirical, research-based approach in building our model.

While we have personality experts on our team, we made the decision that data - and data analytics - would always have the last say in how we defined the model.

Whenever the data showed conclusively that we could include something in our model, we would include it.

Whenever we had an intuition about what people in specific personality types would be like, if we did not have data to support that intuition, we would not include it in the model… or we would gather new data to determine whether or not the intuition matched that expanded data.

Indeed, over the course of building the model, many of our intuitions were flat-out wrong. The data also did not support many commonly-accepted views about personality.

4.1. The Dilemma: Number of Personality Types

A fundamental decision in developing a personality type model is how many personality types to include in the model.

Jung proposed four personality traits, which led to Myers and Briggs developing a model with $2^4 = 16$ personality types. Almost all other well-known personality type systems have followed this same basic pattern of using four traits to define 16 personality types.

We decided to challenge and test this unwritten assumption that a 16-type model was best.

There are obvious benefits to having a small number of personality types in a model:

- It's easy for people to remember and internalize a small number of types
- It's easier for the developers of the system to write the descriptions of each type, as there's simply less analysis and writing to do

But this comes at the cost of reducing the number of traits the model takes into account, and so providing a *less comprehensive* model of human personality.

There are also potential benefits to having a larger number of types:

- By including more personality factors in the personality model, there is the potential to create a more comprehensive personality type model
- With the right data analytics this more-comprehensive model could also potentially be more accurate, as people would not be grouped into large pools, but smaller groups with more distinctive statistics about their characteristics.

But this comes at the cost of the model being more *complex* and thus more difficult for people to remember and internalize.

In summary, there is a fundamental dilemma between creating a simple personality type model with a few number of personality types and a comprehensive personality type model with a large number of personality types.

With this dilemma in mind, we set out to create a personality type model that was simple enough to be comprehensible, but comprehensive enough to be useful and provide new insights.

4.2. Data Collection

We built an automated online research platform for data gathering, and gathered data between April 2019 and April 2020.

The research site invited users to take a Big Five personality test, provide their age and gender, and then continue to answer *additional* output descriptions after they had taken the Big Five test. The process of developing these additional descriptions is covered in section 4.2.1. The users rated their agreement or disagreement with each test question and additional output descriptions on a Likert scale of -3 to +3.

Responses to the Big Five test questions were mandatory, and responses to the additional output descriptions were optional. We made this decision as some descriptions were not relevant to all people. For example, descriptions relating to a person's relationship with their spouse are not relevant to people who are not in a committed relationship.

In our preliminary phase of research we used a 100-question Big Five test, which calculates respondents' major Big Five traits but does not calculate scores for their 30 sub-facets. We later transitioned to our final 120-question test, which calculates respondents' Big Five traits plus their 30 sub-facets.

Over the course of 13 months we surveyed a total of 144,164 individuals who completed the Big Five test questionnaires. Of those, 51,925 completed the 100-question test and an additional 92,239 completed our final 120-question test. We only used the data from the 92,239 people who completed the 120-question test in developing our final model.

We also received a total of 16,181,276 ratings of output responses to 9,394 output descriptions.

From a user's answers to 120 Big Five test questions, we calculated their trait scores for the Big Five traits and scores for the 30 facets using the scoring scale provided by Johnson. The traits and facets are calculated as a real number between -1 to +1 with one decimal place.

In our data analytics we take each user's 30 sub-facet scores plus age and gender as inputs (independent variables) and their responses to the additional descriptions as outputs (dependent variables).

In the remainder of this paper we refer to a user's major Big Five trait scores and their 30 sub-facet scores collectively as their "trait scores." We will refer to a user's rating of additional questions as their output description scores.

4.2.1. Output Descriptions

As we aimed to produce as comprehensive a model of personality as possible, we sought to include as many output descriptions, covering as wide a range of areas of psychology, beliefs and behaviors, as we could practically assemble.

The public domain materials on The International Personality Item Pool (IPIP) site provided a valuable resource. We curated 4,452 descriptions from that site ranging from topics like "Achievement-Striving" and "Adaptability" to "Warmth", "Wisdom" and "Workaholism".

We wrote additional descriptions ourselves, based on our behavioral experts' experience working in corporate training and personal development settings, and a general sense of what we wanted to include in our model. For example, we wrote large numbers of descriptions relating to drivers (motivators in life), emotions (emotional affect and preferences for emotional connection), relationship beliefs and dynamics, lifestyle preferences, decision making styles, conflict and communication styles, life goals and life outcomes.

As noted earlier, when our data analytics suggested gaps between our intuition of people's personality types and the data we had to support those intuitions, we wrote and tested additional output descriptions. If the additional data supported our intuitions, we would include the results in the model and if not, the intuitive hypothesis would be rejected.

4.2.2. Demographics

All of our research was conducted in Spanish, and the vast majority of respondents to our survey were in Latin America. 67.2% of respondents were female and 32.8% of respondents were male. The average age of respondents was 32. There were more respondents in their thirties than any other decade bin, with good representation from 20 through 80 and only a small number of respondents over 80.

When we performed our analytics, we took random samples from our dataset of equal numbers of men and women, so that the analytics would not be gender-biased because of the gender sample bias.

5. APPROACHES TO DEVELOP THE MODEL

We explored four distinct approaches to building our personality model:

1. Classification based on the Big Five traits (see section 5.1)
2. Partitioning the Trait Space by Clusters of Users (see section 5.2)
3. Classification based on output description values (see section 5.3)
4. Grouping Output Descriptions into Themes (see section 5.5)

The first three approaches all failed to produce useful results, but contributed to our understanding of the problem space and allowed us to discover the last approach. The last approach succeeded and led to the development of our current personality model.

5.1. Classification Based on Big Five Traits

As noted earlier, existing personality type models define personality types as combinations of - and + for each trait. For example, the Myers-Brigg type INTJ has:

- +1 on Introversion
- +1 on Intuition
- +1 on Thinking
- -1 on Prospecting (i.e. Judging)

Our first hypothesis was that the same classification approach, but using the five major traits of the Big Five instead of the four traits used in most other personality type systems to classify users into personality types, would generate a useful personality model.

The thinking was that if four trait models led to a useful classification model to create personality types, then given the better technical characteristics of the Big Five traits, and the use of five traits instead of four could give a more accurate and useful personality type model.

This would lead to $2^5 = 32$ personality types.

Using the data we had collected to date, we assigned the users who had completed the test into these 32 putative personality types. We calculated the mean and standard deviations of actual output responses for all users in each of these 32 putative personality types.

We found two problems with this approach:

1. The statistical characteristics of response scores to output descriptions were poor.
2. The putative types did not provide useful insight into human personality

5.1.1. Poor Statistical Characteristics

For a 'good' type model, there should be descriptions that, on average, score high (or low) for people in one type, with low standard deviation (i.e. high predictability), and score low (or high) for people not in the type.

Apart from the descriptions we tested that were essentially re-statements of the Big Five survey questions, we found very few output descriptions that met these criteria, and far too few to build any meaningful type model.

5.1.2. No Useful Insight

Even for the few descriptions that did meet our statistical criteria of being 'descriptive' of the putative type, there were no discernable patterns that made sense in terms of expanding understanding of human psychology or behavior, beyond just restating the Big Five trait descriptions.

They gave no insight into how people think and behave in our key areas of interest: romantic relationships, career path, lifestyle and working relationships.

Thus we eliminated this approach from further investigation, and started exploring more sophisticated approaches to developing our personality model.

The obvious next step was to explore developing a model using the 30 facets of the Big Five model.

5.2. Partitioning The Trait Space By Clusters of Users

The next hypothesis we explored related to finding clusters of users in the 30-dimensional space of 30 facets.

Intuitively, laypeople and psychologists believe that there are common 'types' of people. Or put differently, that there are certain sets of personality traits that tend to appear together.

If this approach played out in terms of the 30-dimensional space of facets, then certain regions of this space would be densely populated with people (representing common groupings of facets that occur together) and other regions would be sparsely populated (representing uncommon groupings of facets that rarely occur together).

If this uneven distribution hypothesis turned out to be true, we could then *define* personality types as the set of these densely populated regions, and *describe* the personality types based on the mean scores of the output descriptions for people in those various regions.

As mentioned earlier, Gerlach *et al* [21] adopted this approach to find four clusters of people in the 5-dimensional space of the Big Five major traits, so there was a precedent for this being a useful approach.

While Gerlach's work found four rather generic personality types, it was still plausible that looking for clusters of people in the 30-dimensional facet space could find more interesting and distinctive clusters.

However, there is a problem that comes with increasing the number of dimensions when searching for clusters in a higher-dimensional space. This relates to the exponential growth of the volume of the space as the number of dimensions increases.

To illustrate this problem, in a 5-dimensional space, there are $2^5 = 32$ quadrants of high and low scores, formed by dividing the trait space into the 32 combinations of high and low scores.

In a 30-dimensional space, there are $2^{30} = 1{,}073{,}741{,}824$ quadrants of high and low scores, formed by dividing the trait space into the 1,073,741,824 combinations of high and low scores.

Thus, with the same number of samples, a 30-dimensional space is 33,554,432 times more sparsely populated than a 5-dimensional space. As a result, the fact that Gerlach *et al* could find clusters in a 5-dimensional space was no guarantee that there would be meaningful clusters in the 30-dimensional space.

For this approach to pan out, there would have to be a very uneven distribution of users, with some very densely-populated regions and a huge volume of very sparsely populated regions.

5.2.1. Distribution of People Across Quadrants
For our analysis we asked three questions:

- Are all of the possible quadrants actually populated?
- Are there some quadrants that are densely populated, and thus important?
- Are there some quadrants that have few to no people in them, and so can be practically ignored?

At the time we performed this analysis, we had surveyed about 14,000 people.

Looking at the distribution of people across the quadrants we found that out of $2^{30} = 1{,}073{,}741{,}824$ quadrants defined by the 30 traits, 13,170 quadrants were populated. Of those, only 582, or 4.4%, contained more than 1 person. The remaining 95.6% of quadrants only contained one person.

Thus we concluded that in the 30-trait space, quadrants do not display any strong patterns of a few quadrants being highly populated and most quadrants being empty or almost empty. To the contrary, the distribution of people across quadrants is fairly even and very sparse.

We also tested clustering adjacent quadrants (i.e. combining neighboring quadrants in one or more trait dimensions) to see if there were patterns of more densely populated regions. We found no densely-populated regions using this approach either.

Thus we found the hypothesis that there were densely-populated regions of the trait space to be invalid.

If there truly are common patterns of personality types, then they do not relate to common patterns in the density distributions of people in the trait space.

5.3. Classification Based on Output Scores
Our next hypothesis was that we could cluster users into personality types based on common characteristics of their scores for the output descriptions.

This hypothesis gives up the assumption that personality types are defined by dividing people into groups based on high or low scores in their facets, and instead looks for regions in the trait space where output scores are similar.

This is a subtle but important distinction, so we will elaborate on this.

All existing personality models we are aware of make the assumption that patterns in trait scores lead to patterns in personality types, and so the geometry of personality types are defined by a simple and direct relationship from their trait scores to personality type. With this approach, 'types' are cubes in the trait space.

Instead, we would be using machine learning to find arbitrary-shaped regions where there were similar patterns of output scores, without the assumption that these regions would be cubes or combinations of cubes in the trait space.

Importantly, these algorithms could now potentially find 'midline' personality types for people who have midline trait scores in some dimensions. However, exploring a higher dimensional trait space for patterns creates a fundamental technical problem, known as "the curse of dimensionality."

5.3.1. The Curse of Dimensionality

As the number of factors included in developing a model increases, the amount of data required to get statistically-meaningful results increases exponentially.

> The common theme of [higher dimensional] problems is that when the dimensionality increases, the volume of the space increases so fast that the available data become sparse. This sparsity is problematic for any method that requires statistical significance. In order to obtain a statistically sound and reliable result, the amount of data needed to support the result often grows exponentially with the dimensionality. Also, organizing and searching data often relies on detecting areas where objects form groups with similar properties; in high dimensional data, however, all objects appear to be sparse and dissimilar in many ways, which prevents common data organization strategies from being efficient. [22]

Using machine learning with a four or five trait model is tractable in terms of data sample requirements, but working with the 30 facets of the Big Five gives over 1 billion quadrants from just dividing each trait into dichotomous high and low scores. Adding age and gender as additional trait dimensions increases this to over 4 billion quadrants.

To get only 100 user samples in each of these 4 billion quadrants, we would need to survey 100 billion individual people. That's 12 times the current world population and roughly equal to the total number of humans who have ever lived.

Thus a simple sampling approach would not be viable for building a model taking 32 traits as input and mapping those to a personality type model.

We overcame this problem by building linear regression models.

5.3.1.1. Linear Regression

We used the ggplot2 package in the R language to generate plots of one-dimensional slices of individual facet dimensions plotted against observed output responses. This allowed us to visualize the relationships between facets and output responses for all the output responses we had tested to date.

We visually inspected 15,610 such plots and found that the vast majority of descriptions displayed roughly-linear relationships between the facet score and output response score in each facet dimension.

Figures 1-4 were generated with the geom_smooth function with method = 'loess', and the smoothing coefficient, 'span' set to 1. The gray bands around the smoothed plots show the 95% confidence interval. The x-axis represents the facet score on a scale of -1 to +1 and the y-axis represents the average observed values for the output description in question. In our data processing the Big Five facets were labeled as input_1 through input_30. The output descriptions were labelled X1, X2… in sequential order according to when we added them to our test environment.

Based on these visual observations, we decided to investigate using multiple linear regression models to predict output responses for users based on their trait scores.

> In statistics, **linear regression** is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one

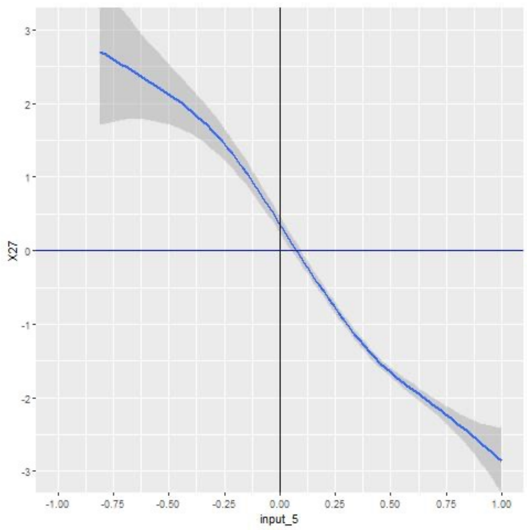

Figure 1: Big Five facets (input_5) versus output (X27)

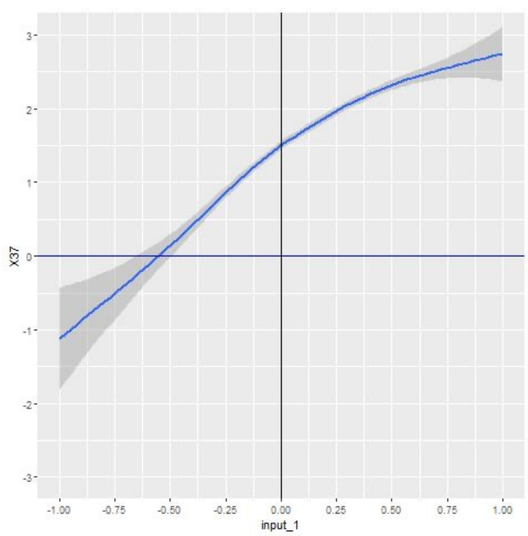

Figure 2: Big Five facets (input_1) versus output (X37)

explanatory variable, the process is called **multiple linear regression**. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

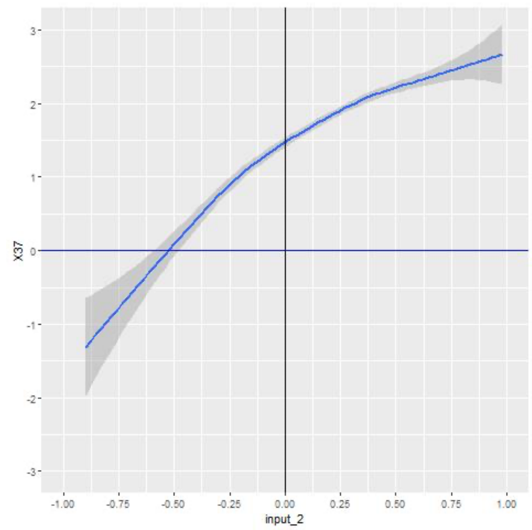

Figure 3: Big Five facets (input_2) versus output (X37)

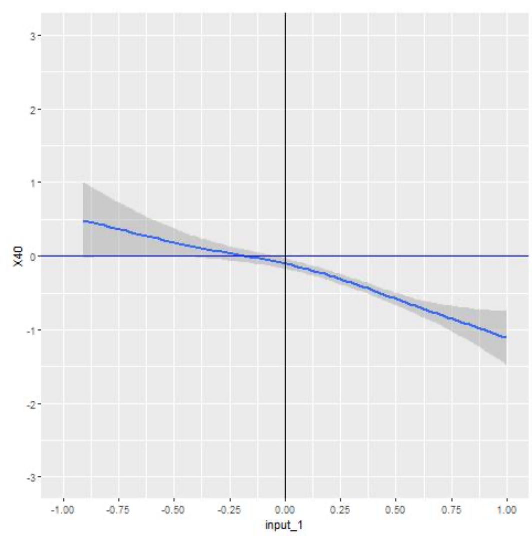

Figure 4: Big Five facets (input_1) versus output (X40)

Using linear regression to predict outputs based on personality traits is not a novel approach, for

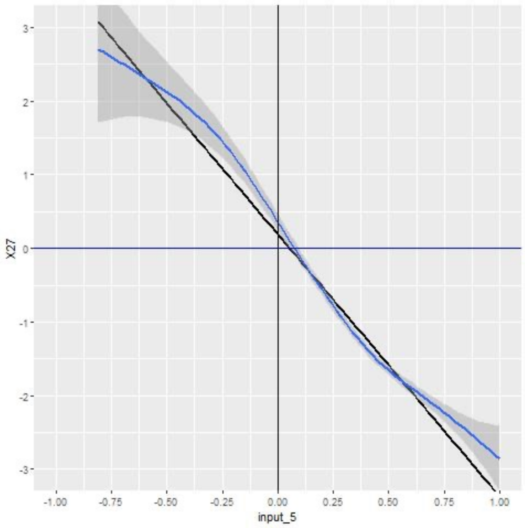

Figure 5: Linear regression of Big Five facets (input_5) versus output (X27)

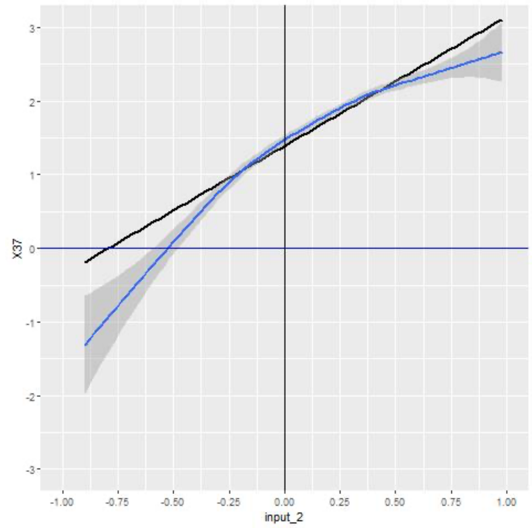

Figure 7: Linear regression of Big Five facets (input_2) versus output (X37)

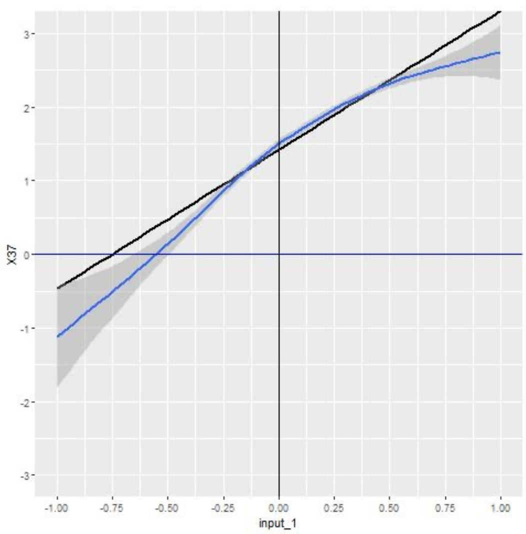

Figure 6: Linear regression of Big Five facets (input_1) versus output (X37)

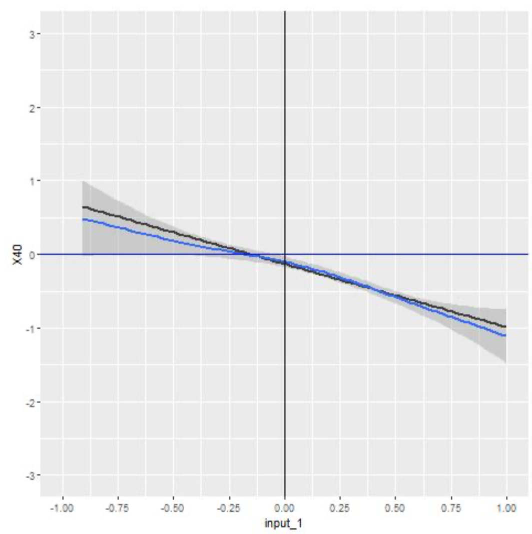

Figure 8: Linear regression of Big Five facets (input_1) versus output (X40)

example, see Vaughn *et al* [23] and Land and Manner [24].

Figures 5-8 show the same plots as Figures 1-4, with the linear regression model superimposed (black line) on the LOESS smoothed functions.

Using standard linear regression methods we achieved R^2 values between 0.3 and 0.4 across all of the descriptions we modeled. This is in line with typical R^2 values achieved in psychometrics.

By empirical trial and error, we found that once we reached 1,000 samples for a given output description, the R^2 values did not increase significantly and the regression coefficients did not change significantly as we added additional samples. Thus we established a minimum sample of 1,000 data points to generate linear regression

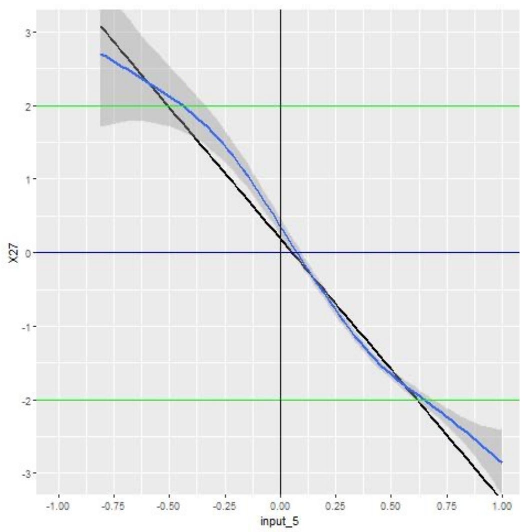

Figure 9: Linear regression of Big Five facets (input_5) versus output (X27)

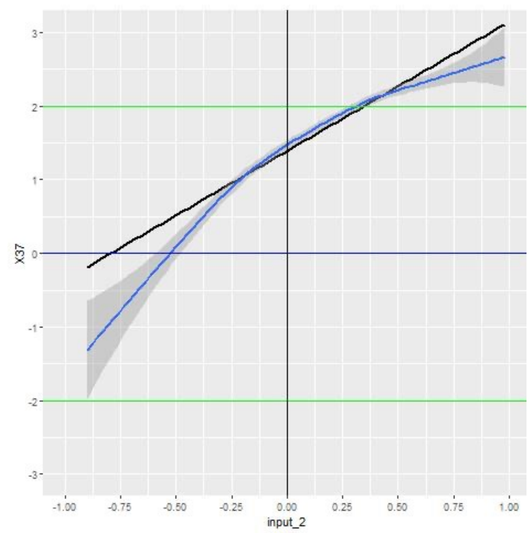

Figure 10: Linear regression of Big Five facets (input_2) versus output (X37)

models for every output description we use in our model.

In our research, we are interested in being able to predict which people will give extreme high or low scores to a description, as these say something distinctive, and so interesting, about a person. Thus, predicting accurately who will give a score $\leq$ -2 or $\geq$ +2 to an output description is more important than accurately predicting the score of people who would give a description a moderate score in the range - 2 to +2.

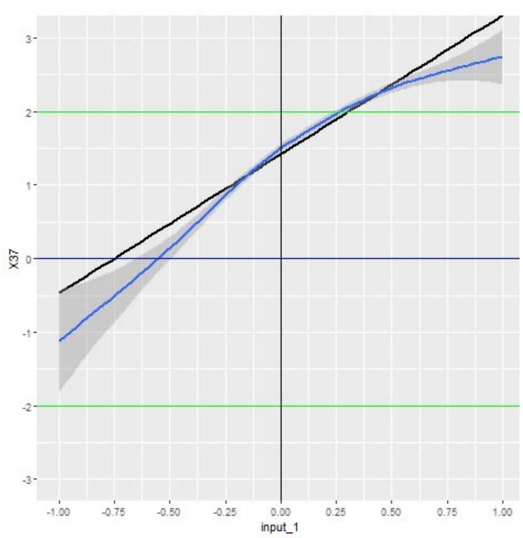

Figure 11: Linear regression of Big Five facets (input_1)

We found that a fairly high proportion of plots have an s-shape. This means that the practical predictive power of the linear regression functions for predicting scores $\leq$ -2 or $\geq$ +2 are often much better than the R^2 values would indicate, as at the extremes, the linear regression predictions are more conservative than the LOESS functions. Thus when the linear regression function predicts that a person will have an extreme high or low score, the LOESS confidence bounds are generally greater than 95%. Thus the predictive power of the linear regression functions at the extremes is significantly better than the R^2 values taken out of the context of how we use the predictions would indicate.

This is illustrated in the plots in Figures 9-11, with the horizontal green lines showing the extremes of -2 and +2.

5.3.1.2. MARS Modeling

In an attempt to achieve higher R^2 values, we explored using Multivariate Adaptive Regression Splines [25] to model our data.

We found that we needed a much higher sample size to get accurate MARS models. Moreover, even with 10,000 samples, MARS did not produce significantly higher R^2 values than simple linear regression. Also, MARS models are much more difficult to interpret and use in other algorithms.

Given that MARS models increased complexity and sample size requirements without significantly improving predictive power, we did not pursue them further.

5.3.1.3. Using Linear Regression to Overcome The Curse of Dimensionality

Our linear regression models allow us to predict the response of a given user to an output description even if we have never sampled a user with the same or even similar trait scores.

Thus, the linear regression models allowed us to overcome the curse of dimensionality. We relied on the predicted values from linear regression models as proxies for actual sampled data in subsequent data analysis. That said, we always gathered at least 1,000 samples for a given output description before using its regression function in our data analysis, as we had found that sample size was sufficient to create a near-optimal regression function.

With the curse of dimensionality tamed, we could continue to explore classification based on output values in the 32-dimensional trait space.

5.3.2. Classification Based on Output Values

We used standard machine learning approaches to find and cluster regions in the 32-dimensional trait space where users had similar predicted output scores for sets of output descriptions.

All of these approaches produced similar results: as the clustering runs progressed, they produced less and less distinctive clusters (or putative personality types) with poorer and poorer differentiation between people in each cluster and outside each cluster.

In other words, as the number of putative personality types was decreased, the quality of the personality types degraded.

Using various clustering approaches, we could find no reasonably small number of personality types that were distinctive and useful.

> Only one pattern remained clear and distinctive as the number of clusters was reduced: There is one set of people who are generally confident and successful and another set of people who are generally anxious and unsuccessful. (We use the generic terms of "confident and successful" contrasted with "anxious and unsuccessful" intentionally. There were hundreds of descriptions in each of these clusters, so more specific labels would be misleading.)
>
> While the finding that this was the most dominant differentiator between sets of people is interesting and useful, having a two personality type model with one type of confident and successful people and a second type of anxious and unsuccessful people failed to meet our criteria for developing a new and useful personality type model.
>
> Having said that, it is notable that none of the well-known, pre-existing personality type models address this fundamental finding, which, according to our research is the most dominant and predictable pattern of human psychology based on the Big Five traits.
>
> In the model we eventually developed, this distinction is the primary factor in our "Reassurance/Outcomes" theme.

Analysis of the results from exploring multiple machine learning approaches led to a simple, key insight:

> *There is no statistically-valid personality type model that is both simple and comprehensive.*

5.4. Simple or Comprehensive

By applying various machine learning approaches, we discovered that it's fairly easy to create a statistically valid personality type model when only a small number of traits, or areas of human psychology and behavior, are taken into account.

For example, when we limited our analysis to descriptions about confidence, fear, success and failure, it was easy to create a very strong, two-type model that described those factors very well as distinct personality types.

Similarly, when we limited our analysis to descriptions about introversion and extraversion, it was easy to create a very strong, three-type model that described those factors as personality types very well: Extraverted, Ambiverted and Introverted.

Through experimentation, we discovered that there were some combinations of output descriptions that did cluster well together, producing meaningful personality types with strong, distinctive statistics differentiating the types, and other combinations that did not.

> Moreover, we found that there are many factors that human intuition and traditional type models include together in a single personality type that have much better statistical characteristics when separated out into different themes. For instance, it's often said that introverts are creative daydreamers. That may be true of some introverts, but it's not true of all introverts. When the themes are combined, the resulting statistical characteristics, and thus what we can accurately say about people in the theme are poor, and separating them gives more accurate and distinctive descriptions.
>
> As we developed our own model we discovered many things that we thought would go together based on our intuition of human personality, but found were in fact better treated as independent factors according to the data. We include an anecdotal example to illustrate this point.
>
> One of our team members (by her own account and our experience of working with her) *hates* having to follow rules and being told what to do. Consistent with this, she falls into Rules Style 1 in Our Model [26].
>
> She also falls into Goals & Energy Style 4. Our description of that style states, "They do better in structured environments where there is clear guidance on what to do when. This structure gives them motivation. They know what to do and can focus on doing it. It's hard for them to be motivated when it's not clear what they should be doing."
>
> In other words, here was someone who hates rules and structure but appreciates being given very clear guidance.
>
> Our behavioral experts found it baffling that a person could fall into both of these seemingly-contradictory styles, so much so that it made us question the accuracy of our model.
>
> So we asked her about it.
>
> To her, the answer was completely obvious and made sense without a moment's thought. "Of course! If you try to *tell* me what to do, I'm going to get angry and rebel. But I often don't know how to approach new projects or know how to get started. So if I *ask* for help and you can give me a clear, step-by-step process, then that's extremely helpful for me."
>
> Time and again, we found that our data analysis was more accurate than human intuition.

So, by trying to combine different themes into a single model, most pre-existing personality models lose statistical accuracy and make assertions that are not supported by the data.

By the time we reached this conclusion, we had gathered sample data on 9,395 output descriptions, covering a wide range of themes.

We now had a new challenge: to separate these 9,395 output descriptions into distinct themes that would have good clustering and statistical characteristics within each theme, while making sense in terms of human behavior and personality psychology.

5.5. Grouping Output Descriptions into Themes

Our final hypothesis was that we could group our 9,395 output descriptions into separate themes where:

- The descriptions in each theme made sense in terms of describing a meaningful aspect of personality.
- Within each theme, the machine learning algorithms could find clusters that had strong statistical characteristics. Specifically, that there would be high mean values for some subsets of output descriptions within each cluster, and high differences of mean values of those subsets of output descriptions compared to other clusters.
- Together the themes would give a comprehensive description of human personalities and types.

With this model of separate themes and clusters within themes, it no longer makes sense to refer to a person's single "personality type."

For clarity, we will refer to clusters within themes not as "personality types" within the theme, but instead as "styles" within the theme. We will refer to a person's entire personality type as the combination of their styles in each theme.

To separate out our 9,395 output descriptions into useful themes, we went through an iterative, interactive process to define and refine the themes:

1. We sorted the descriptions into separate themes that made sense to us in human terms, relying on our intuition based on our expertise and experience in human personality.
2. We ran machine learning algorithms and analyzed results to test whether the themes produced statistically meaningful and consistent clusters of styles within each theme.
3. We used these results to identify which output descriptions did not seem to belong in their current theme, given the statistical characteristics of the clustering results.
4. Based on this new information, we updated our understanding and intuitions about the meanings of each theme, and used this to guide the next iteration of re-sorting the descriptions into themes to produce better results.

From the human perspective, the criteria for grouping descriptions into a theme were:

- All the descriptions address a **meaningful** area of human personality, for example, how people handle conflict, how people relate to emotional connection with others or what people's preferences are for social interaction.
- Together, the descriptions in a theme give **comprehensive** coverage of the theme in question.

From the statistical perspective, the criteria for grouping descriptions into a theme were:

- All the descriptions in a theme have **consistent patterns** in how they cluster together for different users.
- The clusters generated have strong **statistical validity**, in terms of having descriptions that have high mean scores within some clusters contrasted with low mean scores in other clusters.

Oftentimes, based on our personality expertise, we would group output descriptions together that made sense to us given our qualitative experience from working with people, only to find that from the statistical perspective, some of the descriptions did not belong in the theme, as they clustered very differently from other descriptions in the theme.

Thus we went through an exhaustive, iterative process of grouping and regrouping our 9,395 descriptions into different themes to produce a final set that met both the criteria of being meaningful to our behavioral experts and met our statistical criteria.

At each iteration of re-grouping output descriptions into themes, we used our behavioral expertise to give starting estimates of where we thought boundaries between the styles in each theme might be, and then the algorithms recursively adjusted the boundaries to find the shapes and boundaries of personality styles that optimized the statistical qualities of each style.

We continued this iterative process until we developed our final set of 14 themes that meet both our human and statistical criteria (Figure 12-13).

> At each iteration, for each theme we visually inspected machine learning outputs in Microsoft Excel to inform the next round of re-grouping clusters of descriptions into new themes.
>
> This iterative process - combining both machine learning and human analysis - took about 1,000 hours of human time and about 15,000 hours of computer processing time, running mostly on Amazon AWS EC2 z1d.large instances, which have computing power equivalent to approximately twelve 1.0 – 1.2 GHz 2007 Intel Xeon processors and 16GB of real memory.

Sometimes the optimum boundaries *somewhat* matched what we anticipated based on our behavioral expertise, but more often than not, the shape and content of the final optimized styles were very different from what we expected. The discussion of the Extraversion theme in section 5.6.2.1 is one example of this.

5.6. The Final Model

This process produced 14 themes, and within each theme there are between two and eight styles.

In the themes where there are only two styles, the boundary in the trait space between the two styles is quite simple: it is a 31-dimensional (sloped)

Figure 12: Sample of Themes

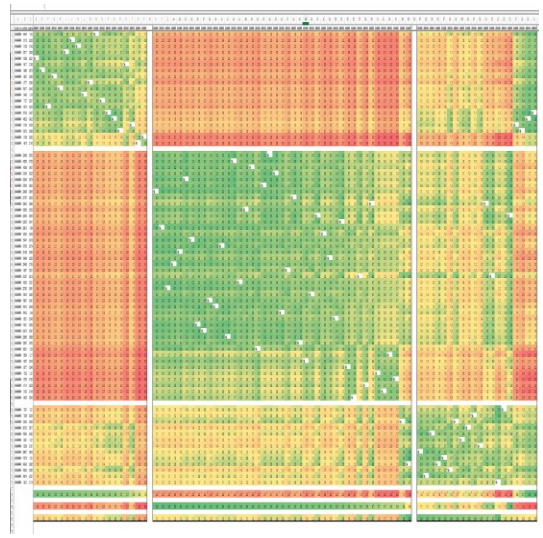

Figure 13: Sample of Clustering

plane dividing the 32-dimensional trait space into two halves.

Importantly, we found that our machine learning algorithms could correctly identify 'midline' personality styles, such as the Ambiverted style in the Introverted-Extravert spectrum.

When the theme contains more than two styles, the geometry of the boundaries between styles is more complex. If a theme has n styles, each style occupies a 32-dimensional polygonal volume in the trait space, bounded by up to n - 1 31-dimensional (sloped) planes.

It's important to note that the shape of these 'midline' styles are more nuanced than just being the average between two extremes. For example, considering the Extraversion theme described in section 5.6.2.1., according to our data, Ambiverts are not just simple averages between Extraverts and Introverts. As discussed in section 5.6.2.1, the three facets that play the most important role in determining which of these three styles a person falls into are Friendliness, Depression and Anxiety.

The following diagram illustrates the boundaries between the three styles in relation to the three facets of Friendliness, Depression and Anxiety. This is an illustration of the relationships rather than an exact plot generated from data.

Geometrically the diagram represents a 3-dimensional cross-section in the three dimensions of Friendliness, Anxiety and Depression through the 32-dimensional space, and then takes a 2-dimensional cross section looking towards the origin point of the three facet axes. This is an illustrative chart rather than an exact plot based on actual data.

It would be possible to develop algorithms to find an appropriate 3-dimensional plane in the 32-dimensional space to take the first cross section, and then find an appropriate 2-dimensional plane to take the second cross section and then generate an exact plot of the boundaries between these styles at that location in the trait space.

However, developing these algorithms would be a non-trivial task and an exact plot would not add anything to the illustrative diagram below. Thus, in this case, we opted to show an illustrative diagram rather than an exact plot (Figure 14).

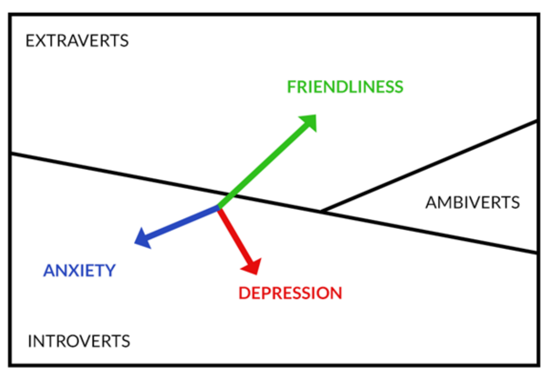

Figure 14: Illustration of Boundaries Between Themes

- The higher a person scores on the Friendliness facet, the more likely they are to fall into the Extravert or Ambivert styles.
- The higher a person scores on the Anxiety facet, the more likely they are to fall into the Extravert or Introvert styles.
- The higher a person scores on the Depression facet, the more likely they are to fall into the Introvert style.

Thus Ambiversion is not just an average between Extraversion and Introversion.

We found that in every theme with three or more styles, each style has its unique and distinctive characteristics, with nuanced relationships between each pair of styles.

5.6.1. The Themes and Styles

In this section we provide some qualitative information about the themes and styles we discovered and built into our current model.

For reference, we provide a summary of each theme and its component styles in the Our Model page on our website [2].

5.6.1.1. Statistical Criteria for Writing Style Descriptions

Once we had identified the 14 themes and their component styles, we needed to write the descriptions for each style.

To do this, we needed to have statistics on the predicted scores people assigned to each style would have for each output description used in the theme.

To generate these data, we developed a machine learning algorithm that performed the following steps:

1. Sample the trait scores of users who had taken our test
2. Predict the output score each user would give to each output description used in the theme (using our linear regression models)
3. Assign each sample user to a style within the theme
4. Assemble summary statistics of the predicted output scores of each output description in the theme for each user assigned to the theme:
    - Mean score of users within the style
    - Standard deviation of scores for users within the style
    - Difference between the mean score of all users within the style and the mean score of all users not in the style

This gave us statistics for writing the summary descriptions for each style, whereby we could ensure that:

- What we said about a style was *accurate*
- What we said about a style was *distinctive* from what we said about other styles
- We had *high confidence* that what we wrote about a style would be accurate for the vast majority of people assigned to the style

As a general guideline, we ignored output descriptions that had low differentiation between styles (thus avoiding the Barun Effect), and wrote the style descriptions based on the 84% confidence level of predicted score (one standard deviation from the mean) for people in the style.

There was one exception to these guidelines. For styles that were truly midline for all of the output descriptions in the theme, we could not write about what was *different* about them, as they were entirely normal or average. In those cases, we wrote for what would be true of them, even though that would also be true for people in other themes. This allowed us to capture and describe truly midline styles in a way that was statistically meaningful and valid.

In writing style descriptions, we were rigorous about not interpolating from the *data* we did have to what our behavioral experts *thought* would be true of people in that style. There were many long discussions relating our behavioral expertise to the data, where our behavioral expertise would say something like, "Well, obviously people in this style would also be like X." and the data would rebut with "If there's no data to support it, we cannot assume it and cannot write it."

[2] https://test.interpersonality.com/data-driven-personality-model

5.6.1.2. Influence of Facets on Style Definitions
In the discussions below, we refer to the 30 facets of the Big Five model. Those facets are defined by the scoring key of the test we use, which you can find here[3].

For each of our styles within each theme, we assembled statistics on the means and standard deviations of facet scores of all people in our sample assigned to the style, plus information about their age and gender. From looking at these mean scores and standard deviations, we can draw some broad conclusions about the relationship between the traits and the style.

The standard deviation across the mean scores of each facet in each style gives an indication of how influential that facet is in defining the theme. When the standard deviation across the mean scores of a facet across styles is high, it indicates that the machine learning algorithm partitions individuals into styles with high dependence on that facet. When that measure is low, it indicates that that facet does not play an important role in assigning an individual to a facet.

In the discussion below, when we list the facets that play the most important role in each theme, they are listed in order of the standard deviation of the mean scores of each facet in each style, highest first, so that we show the most influential facets first.

Having said that, it's important to keep in mind that an individual is assigned to a style based on their entire 30 facet scores plus age and gender, so while some facets play more or less important roles in a theme *in general*, a specific individual's score may differ greatly from the mean scores in the style.

> For example, on average, people in the Assertive style of the Conflict Dynamics theme have higher than average scores for the Depression facet. However, we have a team member who scores very low on the Depression facet but still belongs in the Assertive style - based both on the algorithm that assigns individuals to facets and his own and his colleagues' assessment of his behaviours in regard to this theme.

Thus, while exploring the facets that have most impact on assigning individuals to styles is illustrative of general patterns, it's important to remember that a person's full 30 facet score plus age and gender are used to actually assign individuals to styles - and individual results can vary quite widely from general patterns.

5.6.2. Some Example Themes and Styles
We'll include some illustrative examples below.

5.6.2.1. The Extraversion Theme
Many personality type models include extraversion-introversion as a personality trait in their model. The Big Five model also includes Extraversion as one of its five main traits.

Thus, we fully expected this trait to become part of our model as we curated our output descriptions and developed our themes.

We also expected that the six facets from the Extraversion trait in the Big Five model would have the most important influence on assignment of individuals to the styles in this theme.

However, as our curation process unfolded, we were surprised by the results. The data tell a different story from common perceptions of extraversion-introversion.

The most influential traits for the Extraversion theme are:

1. The Friendliness facet from the Extraversion theme
2. The Depression facet from the Neuroticism theme
3. The Anxiety facet from the Neuroticism theme
4. The Self-Consciousness facet from the Neuroticism theme
5. The Vulnerability facet from the Neuroticism theme
6. The Gregariousness facet from the Extraversion theme
7. The Self-Discipline facet from the Conscientiousness theme
8. The Cheerfulness facet from the Extraversion theme

It would be reasonable to assume that all six facets of the Extraversion theme would play important roles in the Extraversion theme, but this turns out not to be the case. Three of the six extraversion facets turn out to be relatively less important than

[3] https://ipip.ori.org/30FacetNEO-PI-RItems.htm

facets related to Neuroticism and Conscientiousness.

Extraverts and Ambiverts tend to have higher scores for Friendliness than Introverts.

Extraverts and Ambiverts tend to have lower scores for Depression than Introverts. This is consistent with research that has found a correlation between the amount of social time people have in their lives and depression.

Extraverts and Ambiverts tend to have lower scores for Vulnerability than Introverts.

Interestingly, both Extraverts and Introverts tend to score high for Anxiety, while Ambiverts tend to score lower.

Extraverts and Ambiverts tend to score lower for Self-Consciousness than Introverts.

The average scores for Gregariousness follow the pattern one would expect: Extraverts are the most gregarious, Ambiverts are in the middle, and Introverts tend to score low.

Extraverts and Ambiverts score somewhat higher on Self-Discipline than Introverts.

Extraverts and Ambiverts tend to have similarly high scores for Cheerfulness while Introverts tend to have lower scores.

Age also plays a role in this theme. There are more young extraverts and introverts, and as people age, there are more Ambiverts.

The way the different Big Five traits and facets play out in this theme illustrate the benefits of taking a data-driven, machine learning approach to building a personality type model over an approach relying only on human intuition. Human intuition would argue for defining personality styles for Extraverts, Ambiverts and Introverts based only on the Big Five Extraversion trait and its six facets, but it turns out that other traits and facets play equally important roles.

5.6.2.2. The Thinking Style Theme

The Thinking style addresses questions like, "Do you need to have a creative outlet? Do you daydream, playing with abstract ideas, or are you more concrete?"

The most influential traits for the Thinking Style theme are from the "Openness to Experience" trait of the Big Five model:

1. Intellect
2. Imagination
3. Artistic Interests

There's no surprise here, as the theme very closely mirrors those traits.

It is interesting to note that only three of the five facets from "Openness to Experience" play have a primary influence on this style. The Excitement Seeking facet of the Extraversion theme has a more important influence on this style than the other three facets of "Openness to Experience."

5.6.2.3. The Learning Theme

This theme addresses questions such as, "Do you love learning? If you aren't learning something new, do you feel like your 'brain is dying'? Do you hate when people push you?"

The facets that influence this theme are surprising, both on the number of facets that play an important role and the Big Five traits that they come from.

The most influential facet is Friendliness from the Big Five Extraversion trait.

The next four most influential facets are all from the Neuroticism trait:

1. Depression
2. Anxiety
3. Self-Consciousness
4. Vulnerability

The last two important facets are:

1. Gregariousness from the Extraversion trait
2. Self-Discipline from the Conscientiousness trait

This suggests a very rich and complex interplay between the 30 facets and how people relate to learning - on their own and with feedback from others.

In our analysis, we found two styles that enjoy learning:

1. “These people have a thirst to learn. If they’re not learning something new, they may feel like their ‘brain is dying’. They like being in environments that challenge them to get better and better every day.

So whether it's a stack of books, picking up a new skill like cooking, or practicing an existing hobby to get better at it, these people will be happiest when they are learning something new on a regular basis."

2. "These people have high standards and are constantly learning, pushing themselves to be at the cutting edge of anything they do. People are an especially important part of this process for them. Much of their drive is directed towards learning to improve their relationships. For them, people are a source of mutual support and growth."

And a third style that prefers to avoid being given feedback:

3. "These people don't like it when others point out their flaws and push them to change. While they may have plenty of things they would like to improve in their life, they aren't as focused on self development and learning as some."

People in the first two styles generally have higher facet scores for Friendliness, Gregariousness and Cheerfulness.

People in the third style generally have higher facet scores for Depression, Anxiety and Vulnerability.

This suggests that people who are friendly and confident generally welcome feedback and learning more than people who prefer to keep to themselves and are more neurotic.

6. NUMBER OF PERSONALITY TYPES

In our model, we define an individual's "personality type" as the composite of their style assignment within each of the 14 themes.

Thus the total number of personality types in our model is the product of the number of styles in each theme.

This number is 3 x 8 x 4 x 3 x 6 x 5 x 2 x 3 x 2 x 2 x 3 x 2 x 3 x 8 = 29,859,840 personality types, or about 30 million

Thus, we have achieved our goal of having a model that is both meaningful and comprehensible to humans, while still covering the breadth of diversity of human personality types.

By having 14 themes with only two to eight styles within each theme, the model is easy enough for people to understand.

By having about 30 million personality types, the model gives enough differentiation between people to meaningfully describe the vast diversity of human personality types.

7. FUTURE RESEARCH

7.1. Cross-Cultural Validation

We currently generate personality reports in English and Spanish and the test and reports are available online throughout the world in those languages.

As we conducted our research in Spanish and mostly in Latin America, this raises the question of whether the research can be used to draw conclusions for native English speakers from cultures outside of Latin America.

The Big Five personality test has been shown to be robust across a wide range of cultures and languages.

Validity of the test across languages and culture is one issue. There is a separate question of whether the relationship between input variables (test questions responses) and output descriptions (which we used to develop our descriptions of personality types) is stable across languages and cultures.

The data analytics, research and writing of our styles was conducted by a multinational team from the USA, UK and Colombia working in English and Spanish.

We believe that the conclusions of our research will be robust across language and cultures. Small sample tests of sharing profile results with native English speakers from the USA, UK and Europe are consistent with this belief. However, we do not yet have statistically significant sample sizes in English and across non Latin American cultures to validate this claim.

In the near future, we expect to get statistically-meaningful samples from native English speakers around the world, from a variety of cultures. Later we anticipate offering the test and results in additional languages.

As we gather these new data, we will continue to test and reevaluate the results on a language-by-language, region-by-region and country-by-country basis to evaluate whether the results are consistent across demographic groups or whether the model requires adjustment for different demographic groups.

If the data do indicate that any adjustments are required, we will update our model accordingly.

7.2. Supporting the Psychometric Research Community

Our live public site is set up to support the gathering of new research data using a similar process as the one we used on our research site to gather the original data.

After users have completed the 120 Big Five test questions, we have built in the option to invite them to respond to a small number of additional output descriptions, which they can rate on a Likert scale of -3 to +3.

When we have no active research in progress, users will not be shown any additional optional questions. When we do have active research in process, the system will select additional questions to display selected randomly from the pool of currently-active output descriptions we are gathering data on.

Thus, we have a system where we can conduct new research into the relationships between the 30 facets plus age and gender and any output descriptions. We can gather a statistically-valid sample for several hundred output descriptions in a matter of a few days. So, we can achieve fast turn-around on gathering statistically valid samples when we add new output descriptions to test.

We plan to use this platform for our own ongoing research, and invite other respectable researchers to collaborate with us.

We are open to collaborating both with academic researchers who can define their research projects and analyze the data themselves, and with non-academics who have a research interest but would not know how to design the project and conduct the analytics themselves. In the latter case, we would work with the non-academic researcher to define the project, conduct the research, evaluate statistical validity and co-publish the results.

We can test output descriptions provided 'as is' by the other researchers or assist them in developing output descriptions to test based on our experience to date.

For analysis, we can support researchers by applying our existing machine learning algorithms to the data, or export data for researchers to analyze themselves.

Of course, in the case of data export, we fully anonymize all data according to the terms set out in our privacy policy. From a research perspective, the only potential limitation this could create is that geographic regions specified in any exports must contain enough samples to make it impossible to find or guess individual users or small groups of users based on their geographic region.

If you would like to explore conducting a research project using our platform, please email us at research@interpersonality.com outlining your research interest.

Footnotes

"Myers-Briggs", "Myers-Briggs Type Indicator" and "MBTI" are registered trademarks of The Myers & Briggs Foundation.

"DiSC" is a registered trademark of John Wiley & Sons, Inc.

Acknowledgments

We are grateful to contributions of the following organizations and individuals, without which our current research would not have been possible.

International Personality Item Pool

https://ipip.ori.org/

The International Personality Item Pool (IPIP) for curating and sharing over 250 personality trait measurement scales constructed from over 3,000 personality descriptions. The items and scales are in the public domain, which means that one can copy, edit, translate, or use them for any purpose without asking permission and without paying a fee.

Moreover, we thank the many researchers who contributed their items to this site. There are too many to list individually, but they have all made significant contributions to the field and to our research.

Costa & McCrae

Costa & McCrae for developing the first Big Five personality type model and test, The NEO Personality Inventory [5], which has been the foundation for many later developments in the field of psychometrics based on the Big Five model.

John Johnson

John Johnson Professor Emeritus of Psychology, Pennsylvania State University for developing a short Big Five trait test [20] that is both short and accurate, and sharing it in the public domain.

REFERENCES

[1] C. S. Fichten and B. Sunerton, "Popular horoscopes and the 'Barnum effect,'" *The Journal of Psychology*, vol. 114, no. 1, pp. 123–134, 1983.

[2] I. B. Myers, "The Myers-Briggs Type Indicator: Manual," Consulting Psychologists Press, 1962.

[3] S. Christy, "DiSC Assessment Impact on Communication and Understanding of Self and Team," Virginia Tech, 2018.

[4] D. Shaffer, P. Fisher, and C. Lucas, "The diagnostic interview schedule for children (DISC)," John Wiley & Sons Inc, 2004.

[5] P. T. Costa and R. R. McCrae, "Normal personality assessment in clinical practice: The NEO Personality Inventory," *Psychological assessment*, vol. 4, no. 1, p. 5, 1992.

[6] T. L. Bess and R. J. Harvey, "Bimodal score distributions and the Myers-Briggs Type Indicator: fact or artifact?," *Journal of Personality Assessment*, vol. 78, no. 1, pp. 176–186, 2002.

[7] Wikipedia, "Ipsative," *Wikipedia*, 2020. https://en.wikipedia.org/wiki/Ipsative#Psychology (accessed Nov. 01, 2020).

[8] G. D. Chapman, *The five love languages: The secret to love that lasts*. Northfield Publishing, 2015.

[9] O. P. John, R. W. Robins, and L. A. Pervin, *Handbook of personality: Theory and research*. Guilford Press, 2010.

[10] Y. S. Ben-Porath and R. P. Archer, "The MMPI Instruments," Routledge/Taylor & Francis Group, 2014.

[11] U. of Minnesota, "A Brief History of the MMPI Instruments," *University of Minnesota Press*, 2020. https://www.upress.umn.edu/test-division/bibliography/mmpi-history (accessed Nov. 01, 2020).

[12] R. R. McCRAE and P. T. COSTA JR, "The NEO Personality Inventory: Using the Five-Factor ModeI in Counseling," *Journal of Counseling & Development*, vol. 69, no. 4, pp. 367–372, 1991, doi: https://doi.org/10.1002/j.1556-6676.1991.tb01524.x.

[13] P. T. Costa and R. R. McCrae, *Neo personality inventory-revised (NEO PI-R)*. Psychological Assessment Resources Odessa, FL, 1992.

[14] R. R. McCrae and P. T. J. Costa, "Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality.," *Journal of personality*, vol. 57, no. 1, pp. 17–40, Mar. 1989, doi: 10.1111/j.1467-6494.1989.tb00759.x.

[15] R. M. Capraro and M. M. Capraro, "Myers-briggs type indicator score reliability across: Studies a meta-analytic reliability generalization study," *Educational and Psychological Measurement*, vol. 62, no. 4, pp. 590–602, 2002.

[16] R. R. McCrae and P. T. Costa, "Joint factors in self-reports and ratings: Neuroticism, extraversion and openness to experience," *Personality and Individual Differences*, vol. 4, no. 3, pp. 245–255, 1983, doi: https://doi.org/10.1016/0191-8869(83)90146-0.

[17] R. R. McCrae, "Cross-cultural research on the five-factor model of personality," *Online readings in psychology and culture*, vol. 4, no. 4, pp. 1–12, 2002.

[18] J.-P. Rolland, "The cross-cultural generalizability of the Five-Factor model of personality," in *The Five-Factor Model of Personality Across Cultures*, J. McCrae, Robert R., Allik, Ed. Kluwer Academic/Plenum Publishers, 2002, pp. 7–28.

[19] R. R. McCrae, J. Costa Paul T., and T. A. Martin, "The NEO–PI–3: A More Readable Revised NEO Personality Inventory," *Journal of Personality Assessment*, vol. 84, no. 3, pp. 261–270, Jun. 2005, doi: 10.1207/s15327752jpa8403_05.

[20] J. A. Johnson, "Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120," *Journal of Research in Personality*, vol. 51, pp. 78–89, 2014, doi: https://doi.org/10.1016/j.jrp.2014.05.003.

[21] M. Gerlach, B. Farb, W. Revelle, and L. A. Nunes Amaral, "A robust data-driven approach identifies four personality types across four large data sets," *Nature Human Behaviour*, vol. 2, no. 10, pp. 735–742, 2018, doi: 10.1038/s41562-018-0419-z.

[22] Wikipedia, "Curse of dimensionality," *Wikipedia*, 2021. https://en.wikipedia.org/wiki/Curse_of_dimensionality (accessed Nov. 01, 2020).

[23] M. G. Vaughn, M. O. Howard, and M. DeLisi, "Psychopathic personality traits and delinquent careers: An empirical examination," *International Journal of Law and Psychiatry*, vol. 31, no. 5, pp. 407–416, 2008, doi: https://doi.org/10.1016/j.ijlp.2008.08.001.

[24] W. Lane and C. Manner, "The impact of personality traits on smartphone ownership and use," *International Journal of Business and Social Science*, vol. 2, no. 17, 2011.

[25] J. H. Friedman, "Multivariate Adaptive Regression Splines," *The Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 1991, [Online]. Available: http://www.jstor.org/stable/2241837.

[26] Interpersonality Pte. Ltd, "Our Personality Model," *Interpersonality Pte. Ltd*, 2020. https://test.interpersonality.com/our-model (accessed Nov. 01, 2020).