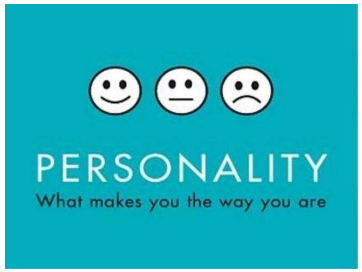
Findings: Astronomical Factors at Birth Make a Significant Difference in Online Personality Self-Assessment and Can Be Useful in Its Classification

by Renay Oshop

Abstract

Using the Big Five Inventory (BFI), a well-established online personality assessment, Renay measured the Big Five personality traits for 1006 different voluntary participants, also collecting the participant's birth date, time and place at the time of assessment. From this birth information, Renay calculated the positions of the Sun, Moon, Mercury, Venus, Mars, Jupiter, Saturn, Uranus, Neptune and Pluto in the International Astronomical Union (IAU) constellations, as well as the planets' prograde or retrograde statuses at birth. All five personality traits changed significantly as planetary positions in the IAU constellations or planetary retrograde status changed. Further, using these IAU planetary placements and retrograde statuses, Renay built an Artificial Intelligence classifier for each personality trait. With the true astronomical data as training input, classifications in extroversion and neuroticism performed significantly better in cross-validation than randomized training input. These results certainly warrant further investigation, seeming to suggest that IAU astronomical factors at birth correlate to online personality self-assessment and can be used to improve classification of an individual as relatively low or high in extroversion or neuroticism.



Introduction

At first glance, astrology and psychology seem to cover much the same ground, both promising comprehensive insight into human personality and behavior.

Renay chose to use neither the Tropical zodiac favored by Western astrologers, nor the sidereal zodiac favored by Vedic astrologers, but rather, the International Astrologer's Union (IAU) zodiac. In this way, self-attribution, of primary concern in astrological research studies of this type, is diminished.

Materials and Methods

Contemporary psychologists have determined through factor analysis the "Big Five" mutually independent, non-overlapping personality traits, of which all others are considered divisions: extroversion, agreeableness, conscientiousness, neuroticism and openness.

The Big Five Inventory (BFI) is a well-regarded clinical instrument that is widely used to assess these Big Five traits (extrovert vs. introvert, etc.) in an individual. An example item for the question BFI is shown in the figure to the right.

Of the 1215 respondents, 209 quiz answers were invalidated because of incomplete birth location or time data or the email used was given for multiple entries with the same birth entry, resulting in 1006 usable submissions.

TAKE THE QUIZ

I SEE MYSELF AS SOMEONE WHO

IS TALKATIVE *

Strongly Agree

Strongly Agree
Agree
Neutral
Disagree
Strongly Disagree

Note that for the whole of the study, no astrology was done, no computing of ascendant, or any other placement. Simply IAU information was used throughout.

The software, Mathematica (v. 11.0), performed all calculations including those for natal astronomy, statistics, and graphing. (Wolfram Research, Inc., 2016) Data files, source code, and some documentation can be found at the online repository . Independent auditing of data and code integrity was performed by Dr. Ivan Stanimirovic.

Stage One

Renay wanted to compare many populations here, two-by-two. For example, what is the median extroversion score for the population whose birth Sun was in IAU Aries, compared to that for whom the birth Sun was in IAU Taurus?

Is it higher or lower? How confident can we be of that assessment given that far fewer than the total n = 1006 are in each group? Now, what about comparing the population with Sun in IAU Aries at birth to the population for whom birth Sun was in IAU Gemini, etc.?

In this study, medians are a more appropriate choice to measure central tendency and hence, similarity of populations, than means, because medians suppress effects of large outliers, thus diminishing the chances for false positives or negatives. Also, they allow easier comparison between sets with non-normal distributions and other skewed distributions for which strict endpoints are not known.

The Mann Whitney test for equivalence of medians for two independent samples with all assumptions verified was chosen to be repeated many times for the first stage whose purpose was to discern between medians for pairs of constellations for Sun, Moon and each planet and between prograde and retrograde for each planet. The popular ANOVA was discounted as not tenable due to its focus on means and its requirements of normality and equality of variances of samples.

All quiz scores were normalized by the POMP method .

The basic process for stage one is the following:

1. For each personality trait (extroversion e.g.):
   - For one astrological indication (Sun IAU constellation at birth e.g.), create pairs (Aries Sun vs. Taurus Sun, Aries Sun vs. Gemini Sun, Taurus Sun vs. Gemini Sun, etc.)
   - Calculate Mann Whitney test for a p-value for equivalency of medians for each pair.
   - Record “true” (indicating significance) if the p-value is equal to or less than 0.05 and "false" if the p-value is greater than 0.05.
   - Tally total of number of “trues.”
   - Divide the number of trues by total pairs possible for that feature. (For example, 13 x 12 = 156 would be the denominator for Sun pairs because Sun has 13 possible IAU constellations.)
   - Store that ratio.

2. Repeat for the other astronomical indications (Moon IAU constellations at birth, Mercury IAU constellations at birth, etc.).

3. Select all ratios not equal to zero. Sort ratios from largest to smallest. Present the actual scores of the largest-ratio astronomical feature for each trait.

Stage Two

Imagine two kiosks, blue and red, side-by-side on your favorite street corner. They both advertise predictive ability about whether you are an extrovert or not, agreeable or not, conscientious or not, neurotic or not, and open-minded or not, as compared to the people that came before you. For either one, all you have to do is type in your real birth date, place, and time. The underlying software has only two steps: it calculates the birth IAU information for you and then uses a machine learning (artificial intelligence or AI) algorithm based only on a database.

For one (the true blue one), the database contains 1006 people's real birth information pointing to their real personality trait profile scores. The database for the other (the random red one) is based on the same 1006 people and the same algorithm, but uses fake birth information, constructed of random rearrangements of the charts, that is pointing to their real personality trait profile scores.

Which kiosk will you go to in hopes of getting the most accurate answer? If you chose the blue one, you may be embracing the possibility of an alternative to a widespread, conventional but as yet uncertain "null" hypothesis that birth data does not matter. Your alternative hypothesis is that, all other things being equal, perhaps it does.

This null hypothesis is a casual yet firm one: astronomical factors have no influence on online personality self-assessment. In other words, a participant with any sort of astronomical chart would still give the same answers to the quiz. We propose as an alternative that the planetary positions and retrogression statuses at birth are indeed indicators to the answers to the quiz. How will you know if the true blue kiosk is indeed the better kiosk? You can compare the kiosk's results to your scores from the BFI answer sheet. Renay is doing the same.

To make things easier for the AI, the scores are separated into high and low for each personality trait, and all Renay asks of it is to tell whether the kiosk visitor is high or low in a trait relative to the focus group the kiosk was trained on. But how high is high? How low is low? There is an answer to this.

Imagine an otherwise empty lecture hall where there are two families represented: one tends to be around 7 feet tall and the other family tends to be around 5 feet tall. There are 3 family members representing each family. If we want to know the two clusters by height of people in the room, it is easy. One could in this special case take an average and count those above the average as one cluster (the 7 footers) and below the average as the other cluster (the 5 footers). But what if the 7 foot family calls in all of its relatives? Let's say there are 20 seven-footers now and still only 3 five-footers. The simplistic solution of taking a total group average no longer works.

There are a variety of ways to derive the proper clusters. For number values (such as with our scores), one method is through Euclidean distance. This is what Renay used.

Renay split each trait into two relative clusters (extrovert vs. introvert, agreeable vs. disagreeable, conscientious versus non-conscientious, neurotic vs. emotionally stable, open vs. closed) through clustering of the BFI scores' Euclidean distances. Using the parameters of these clusters, our participants' responses gave them a membership in one of the two groups for each of the five personality traits relative to the group as a whole being considered. For example, the 5-trait score profile for the first participant is introvert (low in the trait of extroversion), disagreeable (low in the trait of agreeableness), non-conscientious (low in the trait of conscientiousness), neurotic (high in the trait of neuroticism) and open (high in the trait of openness).

The truth of the null hypothesis can be tested by comparing the accuracy of an Artificial Intelligence classifier trained on randomized input (the basis of the null hypothesis) to that trained on the real input (the basis of our alternative hypothesis). If the alternative hypothesis' real chart input performs statistically significantly better in predicting the true status of the participant than the random chart input, then that is support for our alternative hypothesis that the true astronomical charts do matter.

Upon review, we decided to leave out the constellations of the slower-moving planets, Uranus, Neptune and Pluto, since they may be said to also indicate the age of the participant, and age is already known to be an indicator of personality score. The complete 15-feature set of basic astronomical features left after dropping Uranus, Neptune, and Pluto constellations was used to build a machine learning-based classifier for our participants' scores. At the same time, a classifier built on randomized entries was built for each personality trait. A leave-one-out cross validation (LOOCV) was performed on each one, and the accuracies were counted (for example 497 out of 1006). This process was repeated 100 times for the random entries, and the resulting spread of accuracies was analyzed and compared to the accuracy of the real input classifier.

The LOOCV process would be like the following for our kiosks: all but the first volunteer's data is entered. Then, the company assesses, based on the 1005 volunteers' information that the kiosk does know, whether it can accurately predict the first volunteer's personality trait. A 1 is entered in a spreadsheet if the prediction is correct. Otherwise, a 0 is entered.
Then, in a new test, all but the second volunteer's data is entered. This data now includes the first volunteer's. Can the machine predict that second volunteer's personality scores? A 1 is entered for each personality trait it gets right. If the prediction is wrong, a 0 is entered.

In this way, all 1006 volunteers' data are cross-validated, and the spreadsheet shows a column of 1s and 0s. Adding up the 1s for a trait gives a total less than or equal to 1006. The higher the number, the more accurate the machine is for that trait.

Do this process with both the red and blue machines and you have an empirical way to compare accuracies between the two. Similarly, our second stage's purpose was to train an artificial intelligence classifier and test its accuracy compared to random choice input. Some random choices may provide a model that is closer to the truth and some not, so there is apt to be a spread of accuracies as a result.

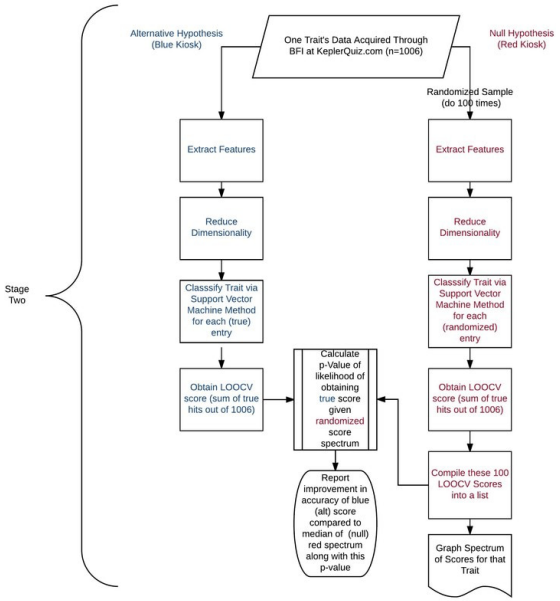
In order to account for the large variety of random shuffling available to the comparison group (which is like the red kiosk) and that spectrum of resultant accuracies, the LOOCV test was repeated 100 times and totals were analyzed to derive parameters for that spectrum or spread. Comparing that accuracy spectrum to the uniform accuracy of the real input engine (which is like the blue kiosk) is the ultimate goal for stage two.

A histogram model was thereby used to acquire a test statistic for each personality trait called the p-value of our alternate hypothesis. This is the equivalent of providing the likelihood of the random input system behaving at least as well as the true input one. Thus, a measure of the likelihood of the alternative hypothesis is acquired.

To interject some common sense, a best guess heuristic was added for comparison where the majority score (the mode) of the training input is chosen as the answer for the trait being tested. For example, for our 1006 volunteers, 526 were classified by clustering to be relatively low in extroversion, *i.e.* introverts. The total count of 1s in the LOOCV should therefore be at least 526, considering that the machine could just count the number of the majority (here, introverts) and still get at least a better than 50-50 score of 526 correct out of 1006. Fisher's Exact Test was used to acquire a p-value of likelihood of getting the modal score or better with a population that matches the 1006 volunteers' median score.

The seven machine learning classifier methods easily available were logistic regression, Markov, naïve Bayes, nearest neighbors, neural network, random forest, and support vector machine. All were tested with the data, and support vector machine performed better than others for accurate predictions of each personality trait with the real input. Thus, this method was also used for the random input classifier for each trait.

The desktop computer that was used for this study was configured with Windows v. 8.1, i7-2600K CPU, and 32 GB RAM. On this computer, computation of the code alone took approximately three weeks.



Flow chart for stage two: comparing classifiers (based on all astronomical features except Uranus, Neptune, and Pluto).

Results

Stage One: Establishing Difference through Pairs of Traits

	Extroversion	Agreeableness	Conscientiousness	Neuroticism	Openness
Placements of pairs with highest ratios of significant differences.	**Jupiter,** Pluto, Neptune, **Sun, Moon, Saturn, Mercury,** Uranus, **Venus, Mars**	Uranus, Neptune, Pluto, **Saturn, Moon, Jupiter, Mars, Mercury**	**Pluto retrograde,** Uranus, Neptune, Pluto, **Sun, Moon, Saturn, Venus, Mercury, Jupiter, Mars**	Pluto, Uranus, Neptune, **Saturn, Venus, Moon, Sun, Jupiter, Mercury**	**Mercury retrograde,** Uranus, Neptune, **Saturn,** Pluto, **Jupiter, Venus, Mercury, Sun, Moon, Mars**

Results for stage one: establishing pairs of traits with statistically significant differences in mean, listed by highest ratio of significant pairs divided by total number of possible pairs to lowest such nonzero ratio.

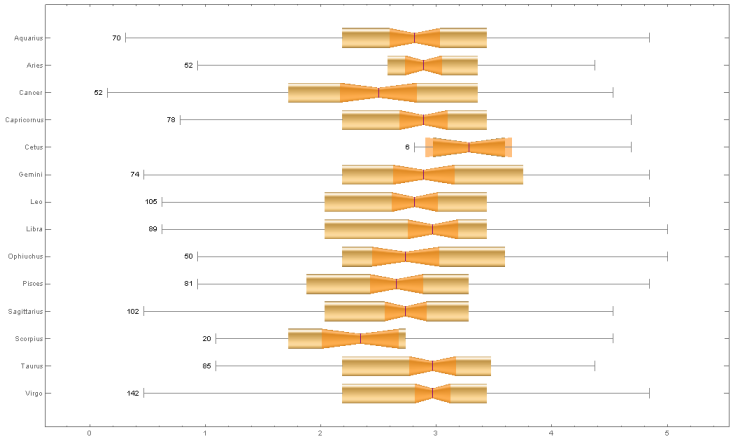
As Uranus, Neptune, and Pluto move slowly through the constellations, they may be said to also be a measure of generations. Thus, while they are hereby acknowledged as potentially essential indicators, they are not included below because any correlation may also be an artifact of the traits being related to the age of the participant, an observation that has already been well-established by conventional psychologists.

Thus, in a move that intentionally ties one hand behind our back and potentially blocks findings, all of the subsequent calculations for stage two did not include Uranus, Neptune, and Pluto natal IAU constellations.

In the following box whisker charts, the x-axis is the POMP-adjust quiz score and the median of the POMP-adjust quiz score is denoted by a purple vertical line. Its bracketed region in orange represents the 95% confidence interval of that median. The bigger region in gold color represents the 25% to 75% interquartile scores for that placement. The thin horizontal line represents the span of the outlier fences of the score values. The number of participants with that particular placement is noted to the left of this thin line.

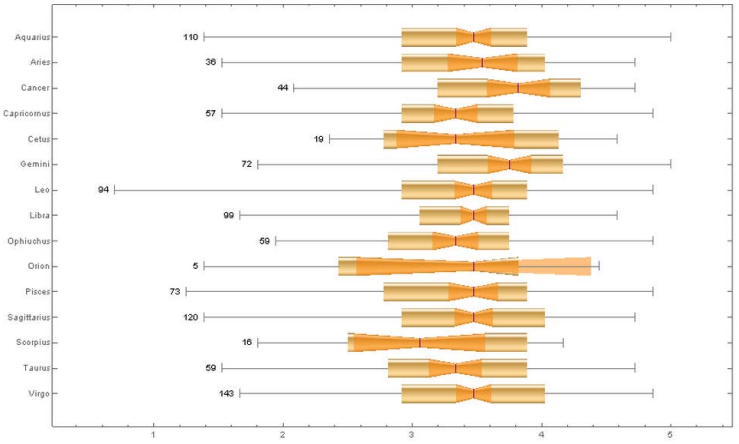
Note that there are 881 total possible pairs of placement for each trait, of which 5% or 44 can be expected to overlap by chance and still be non-significant. Instead, of the 881 pairs: 48, 92, 90, 115, and 88 out of 881 showed non-overlap for extroversion, agreeableness, conscientiousness, neuroticism, and openness, respectively. To save space, for each trait, only the astronomical feature with the highest ratio of non-overlapping placement pairs to total possible pairs is presented in the box whisker charts below.

See below for Jupiter placement differences in the trait of extroversion.



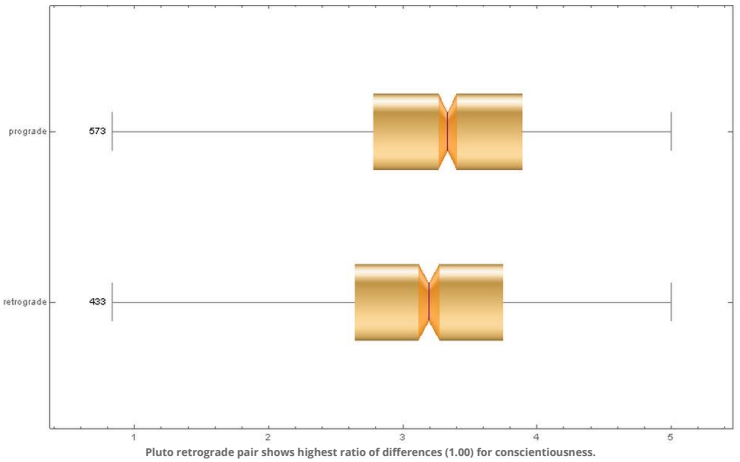
Jupiter placement pairs show highest ratio of differences (0.143) for extroversion.

See below for Saturn placement differences in the trait of agreeableness.



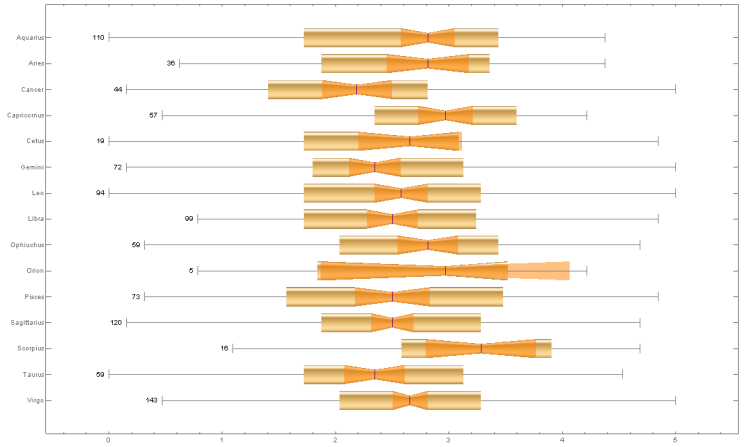
Saturn placement pairs show highest ratio of differences (0.143) for agreeableness.

See below for Pluto retrogression placements for the trait of conscientiousness.



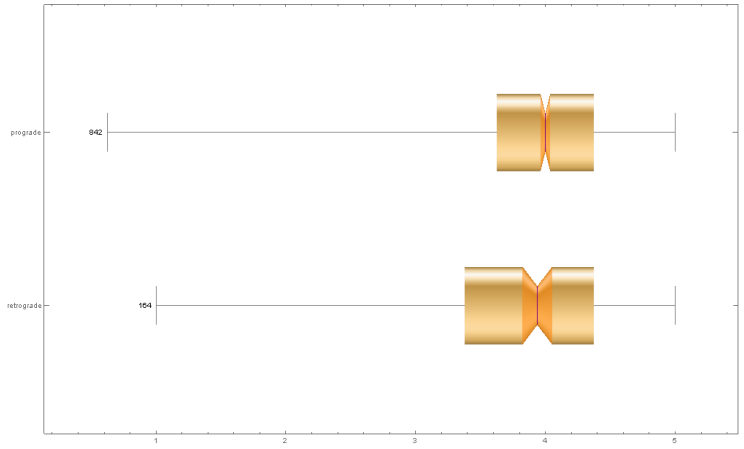
Pluto retrograde pair shows highest ratio of differences (1.00) for conscientiousness.

See below for Saturn placements for the trait of neuroticism.



Saturn placement pairs show highest ratio of differences (0.181) for neuroticism.

See below 9 for Mercury retrogression placements for the trait of openness.



Mercury retrograde placement pair shows highest ratio of differences (1.00) for openness.

Stage Two: Comparing the Accuracy of the Red Kiosk (the Null Hypothesis) to the Blue Kiosk (the Alternative Hypothesis)

See below for data and p-values of performance of the Artificial Intelligence classifiers.

	Extroversion	Agreeableness	Conscientiousness	Neuroticism	Openness
Majority Guess, p-value for Real Input [2,4]	526, p-value < 0.0001	609, p-value = 0.518	586, p-value = 0.518	549, p-value = 0.141	690, p-value =0.519
Median of Random Input Classifier, p-value for Real Input [1, 3, 4]	515.5, p-value = 0.0100	609, p-value=0.750	586, p-value = 0.768	549, p-value = 0.0160	690, p-value = 0.990
Real Input Classifier [3,4]	**650**	**609**	**586**	**574**	**690**
Statistically Significant Difference	Yes, Yes	No, No	No, No	No, Yes	No, No

[1] Based on histogram function of results from 100 trials, 1-sided p-value
[2] Based on Fisher's Exact Test, 1-sided p-value
[3] Uranus, Neptune and Pluto were not included in complete classifiers
[4] Number right out of 1006 total

Stage two results: extroversion shows a significant improvement in the real input compared to both random input and the best guess. Neuroticism classification shows significant improvement using the real input compared to the random input but is not significantly better than the best majority guess, probably because the stricter Fisher Exact Test was used for the latter.

Extroversion is a clear success, but even in considering neuroticism, the Fisher Exact Test for comparing significance of the real input classifier's accuracy to the best guess is considered conservative. For this reason, the significance instead of the real input neuroticism classifier to that for the random input may be said to be most telling, yielding a softer "yes" for success there as well.

Conclusion

Using standard statistical techniques for determining median confidence intervals, it appears that there are many more statistically significant separations between medians of personality trait scores, based on IAU astronomical placements at birth, than can be expected by chance.

As well, building an AI classifier using widely accessible machine learning methods, for the "Big Five" personality traits of extroversion and neuroticism, Renay noticed a statistically significant difference between accuracy of the classifier that was built on true IAU birth placements when compared to a classifier built on randomized data.

This supports the alternative hypothesis that there is some useful information in the IAU placements at birth for classifying participants as low or high in the self-report of extroversion and neuroticism.