

Personality at three levels of analysis

William Revelle

Department of Psychology Northwestern University
Evanston, Illinois USA



NORTHWESTERN
UNIVERSITY

Keynote Address to the 2nd World Conference on Personality
Buzios, Brazil
April, 2016

Slides at <http://personality-project.org/sapa.html>

Partially supported by a grant from the National Science Foundation: SMA-1419324

Outline

Two alternative titles:

1. Personality at three levels of analysis.
 - 1.1 Within individuals
 - 1.2 Between individuals
 - 1.3 Between groups of individuals
2. Personality research: an open and shared science

The importance of open science for personality research can be shown at each of these three levels of analysis.

1. Open source statistics, e.g., The R project
2. Open source materials, e.g., IPIP and ICAR
3. Open source methodology, e.g., Synthetic Aperture Personality Assessment (SAPA)
4. Open source data, e.g., *Journal of Open Psychology Data* and *DataVerse*



Open Science

- Science is an international collaborative endeavor that benefits when more people from more countries participate.
- Scientific societies were started (e.g, the Royal Society in London in 1660) as an “invisible college” to facilitate communication and the sharing of ideas.
- Traditionally we collaborate by publishing our results in scientific journals and by sharing our ideas at national and international conferences or giving guest lectures to our colleagues.
- More recently, there is a trend towards sharing our materials, our methods, and our results, even our data, on the web.
- This makes for better science.



Open Science and the problem of replication

- The last several years has seen a plethora of papers reporting failures to replicate results. This has lead some to worry about the strength of our findings and others to question what does it mean to “replicate” or reproduce a result.
- Others have suggested that we should be more open in our designs, publishing what we plan to do independent of what we actually find.
- This is an important problem that should not be ignored, although pre-registering might inhibit exploratory research.
- But, open science is much more than protecting us from type I errors. It is a philosophy of collaboration. That is what I want to emphasize today.



A Short History of Science: Instrumentation and Modeling

New methods of data collection led to new models and theories.

Instrumentation

1. Telescopes (Galileo)
2. Sailing ships (e.g., Beagle)
3. Depth sounders
4. CO_2 measurement (Keeling)
5. The internet (Al Gore?)
6. WWW (Tim Berners-Lee)
allows for remote assesment
7. Cellphones (e.g., Steve Jobs) allow for repeated mobile assessment

Theories and Models

1. Newton (*Principa*)
2. Darwin/Wallace
3. Plate Tectonics
4. Climate change
5. Open Science
6. Modeling the data across very large sample
7. Modeling within subject variation over time (process models)



A Short History of Science: Part 2: Mathematics and Statistics

1. Calculus (Newton/Leibniz)
2. Data visualization (Playfair to Tukey to Cleveland to Tufte)
3. Probability theory (Fermat/Pascal) and the normal curve (Gauss/Quetelet)
4. Correlation (Galton/Pearson)
5. Factor analysis (Spearman/Thurstone) and Principal Components analysis (Pearson/Hotelling)
6. Discrete (experimental) conditions and the t and F (seeds x manure) distributions (Gossett and Fisher)
7. Main frame computation (Ada Lovelace, John von Neumann, Grace Hopper)
8. Randomization and resampling (empirical distributions, not idealized)
9. Open statistical software for all of us, e.g., R.

We have new methods for collecting, analyzing and sharing our data, lets use them to improve our science.

Four types of openness:

1. Open source software: The R project (R Core Team, 2015)
2. Open source materials:
 - The International Personality Item Pool (IPIP) (Goldberg, 1999)
 - The International Cognitive Ability Resource (ICAR) (Condon & Revelle, 2014)
3. Open source methodology: The Synthetic Aperture Personality Assessment Project (Revelle, Wilt & Rosenthal, 2010; Revelle, Condon, Wilt, French, Brown & Elleman, 2016)
4. Open source data:
 - Data from the ICAR project (Condon & Revelle, 2016, 2015a)
 - Data from SAPA studies (Condon & Revelle, 2015c,b)

In the process of summarizing the last several years of my students and my research, I will show how we use open source software, items, and methods and then share them with the world.

Four types of openness:

1. **Open source software: The R project** (R Core Team, 2015)
2. **Open source materials:**
 - The International Personality Item Pool (IPIP) (Goldberg, 1999)
 - The International Cognitive Ability Resource (ICAR) (Condon & Revelle, 2014)
3. **Open source methodology: The Synthetic Aperture Personality Assessment Project** (Revelle et al., 2010, 2016)
4. **Open source data:**
 - Data from the ICAR project (Condon & Revelle, 2016, 2015a)
 - Data from SAPA studies (Condon & Revelle, 2015c,b)

In the process of summarizing the last several years of research, I will show how we use open source software, items, and methods and then share them with the world.

Part I Open Statistics: R

Part I: Open Statistics: R

R: open source statistical system

What is R

Use R for replications and extensions

Getting and using R

Useful packages

Part II: Open Materials



R: What is it?

1. R: An international collaboration for applied statistical research
 - Originally developed in New Zealand in 1991-93
 - Comprehensive R Archive (CRAN) run out of Vienna
 - Core R members in Austria (2), Canada, Denmark, France, Germany (2), India, New Zealand (3), Switzerland, US (6), UK
2. R: The open source - public domain version of S+
3. R: Written by statisticians (and some of us) for statisticians (and the rest of us)
4. R: Not just a statistics system, also an extensible language.
 - This means that as new statistics are developed they tend to appear in R far sooner than elsewhere.
 - R facilitates asking questions that have not already been asked.
 - “Easy” to learn, harder to master.

Statistical Programs for Psychologists

- General purpose programs
 - R
 - S+
 - SAS
 - SPSS
 - STATA
 - Systat
- Specialized programs
 - Mx
 - EQS
 - AMOS
 - LISREL
 - MPlus
 - Your favorite program

Statistical Programs for Psychologists

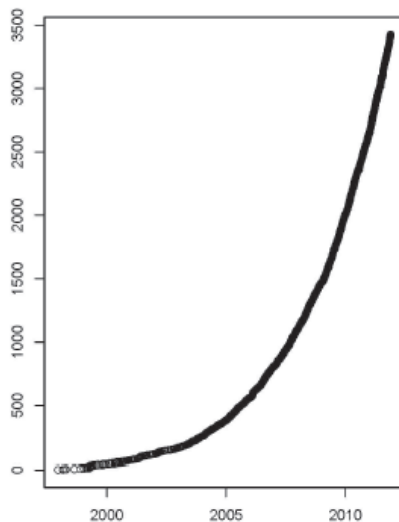
- General purpose programs
 - R
 - \$+
 - \$A\$
 - \$P\$\$
 - \$TATA
 - \$y\$at
- Specialized programs
 - Mx (OpenMx is part of R)
 - EQ\$
 - AMO\$
 - LI\$REL
 - MPlu\$
 - Your favorite program

R: A brief history

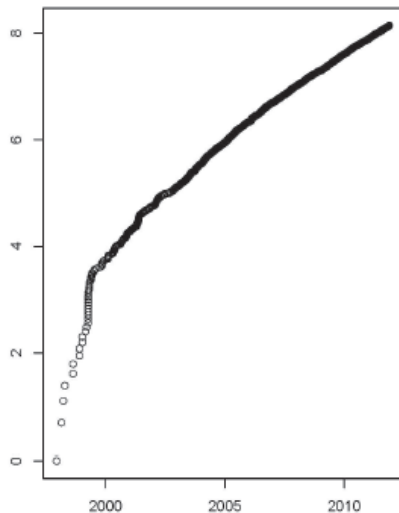
- 1991-93: Ross Ihaka and Robert Gentleman begin work on R project for Macs at U. Auckland (S for Macs).
- 1995: R available by ftp under the General Public License.
- 2001-2016: Core team continues to improve base package with a new release every 6 months (now more like yearly).
- Many others contribute “packages” to supplement the functionality for particular problems.
 - 2003-04-01: 250 packages
 - 2004-10-01: 500 packages
 - 2007-04-12: 1,000 packages
 - 2009-10-04: 2,000 packages
 - 2011-05-12: 3,000 packages
 - 2012-08-27: 4,000 packages
 - 2014-05-16: 5,547 packages (on CRAN) + 824 bioinformatic packages on BioConductor
 - 2015-05-20 6,678 packages (on CRAN) + 1,024 bioinformatic packages + ?,000s on GitHub/R-Forge
 - 2016-03-31 8,182 packages (on CRAN) + 1,104 bioinformatic packages + ?,000s on GitHub/R-Forge (increased by 60 in last 10 days)

Rapid and consistent growth in packages contributed to R

Number of Active CRAN Packages



Log Number of Active CRAN Packages



R as a way of facilitating replicable science

1. R is not just for statisticians, it is for all research oriented psychologists.
2. R scripts are published in psychology journals to show new methods:
 - *Psychological Methods*
 - *Psychological Science*
 - *Journal of Research in Personality*
3. R based data sets are now accompanying journal articles:
 - The *Journal of Research in Personality* now accepts R code and data sets.
 - JRP special issue in R.
 - The replicability project has released its data and R scripts.
4. By sharing our code and data the field can increase the possibility of doing replicable science.

Reproducible Research: Sweave and KnitR

Sweave is a tool that allows to embed the R code for complete data analyses in \LaTeX documents. The purpose is to create dynamic reports, which can be updated automatically if data or analysis change. Instead of inserting a prefabricated graph or table into the report, the master document contains the R code necessary to obtain it. When run through R, all data analysis output (tables, graphs, etc.) is created on the fly and inserted into a final \LaTeX document. The report can be automatically updated if data or analysis change, which allows for truly reproducible research.

Friedrich Leisch (2002). Sweave: Dynamic generation of statistical reports using literate data analysis. I

Supplementary material for journals can be written in Sweave/KnitR so that others can redo or extend the analyses.

What is so great about reproducible research?

1. Allows us to share methods with our collaborators.
2. This can be other labs who want to know what you did. It can be your students, it can even be you.
3. David Condon has suggested that your closest collaborator is you, six months ago, but you don't answer your emails.
4. That is, scripted analyses are for you.
5. The Reproducibility Project (<https://osf.io/ezcuuj/>) has released their 100 replication data set and the R code to analyze it. If any one finds errors or needs more information, they are happy to provide it.
6. See, for instance Dan Gilbert et al. **critique** (Gilbert, King, Pettigrew & Wilson, 2016) and the **response** (Anderson, Bahník, Barnett-Cowan, Bosco, Chandler, Chartier & Cheung, 2016). Also see the recent denial of the need to worry about replication by Roy Baumeister.

Misconception: R is hard to use

1. R doesn't have a GUI (Graphical User Interface)
 - Partly true, many use syntax.
 - Partly not true, GUIs exist (e.g., R Commander, R-Studio).
 - Quasi GUIs for Mac and PCs make syntax writing easier.
2. R syntax is hard to use
 - Not really, unless you think an iPhone is hard to use.
 - Easier to give instructions of 1-4 lines of syntax rather than pictures of menu after menu to pull down.
 - Keep a copy of your syntax, modify it for the next analysis.
3. R is not user friendly: A personological description of R
 - R is Introverted: it will tell you what you want to know if you ask, but not if you don't ask.
 - R is Conscientious: it wants commands to be correct.
 - R is not Agreeable: its error messages are at best cryptic.
 - R is Stable: it does not break down under stress.
 - R is Open: new ideas about statistics are easily developed.

What makes R so powerful are the $> 8,100$ contributed packages

- psych** A general purpose toolkit for psychological research with a particular emphasis upon
- Basic descriptive statistics and basic graphical tools.
 - Basic psychometric procedures including functions for finding $\alpha (= \lambda_3)$, ω_h , and ω_t
 - More advanced data reduction techniques using factor analysis, principal components analysis, and cluster analysis.
 - Introductory Item Response Theory and Multi-level modeling
- lavaan** Basic and advanced structural equation modeling (“The gateway package to R”).
- sem** Structural equation modeling
- lme4** Multilevel modeling.

Short courses and workshops emphasize training in basic and advanced R

1. Symposia

- International Society for Study of Individual Differences (2005)
- 1st World Conference on Personality (2013)
- Society for Personality and Social Psychology (2015)

2. Short courses

- European Conference on Personality (2012, 2014)
- Association for Research on Personality (2012)
- Association for Psychological Science (2013, 2014, 2015, 2016)
- 1st World Conference on Personality (2013)
- STuP 2015 preconference: Confirmatory factor analysis in the lavaan package (R)

3. Summer schools

- Summer school sponsored by ISSID, EAPP and SMEP (2014)

Part II: Open Materials

Part II: Open Materials

Four types of openness:

1. Open source software: The R project (R Core Team, 2015)
2. Open source materials:
 - The International Personality Item Pool (IPIP) (Goldberg, 1999)
 - The International Cognitive Ability Resource (ICAR) (Condon & Revelle, 2014)
3. Open source methodology: The Synthetic Aperture Personality Assessment Project (Revelle et al., 2010, 2016)
4. Open source data:
 - Data from the ICAR project (Condon & Revelle, 2016, 2015a)
 - Data from SAPA studies (Condon & Revelle, 2015c,b)

In the process of summarizing the last several years of research, I will show how we use open source software, items, and methods and then share them with the world.

Part II: Open Materials

Part II: Open Materials

Temperament, Abilities, and Interests: considering appetites and aptitudes

Temperament, Abilities, and Interests: TAI

IPIP: The International Personality Item Pool

Lew Goldberg and the development of the IPIP

Extending the IPIP to include more domains

ICAR: International Cognitive Ability Resource

An international collaboration to measure ability with open source items

Analysis of ICAR items

Part III: Open Methods

Personality, prediction, and life outcomes

1. It has long been known that to predict real world outcomes we need to study more than just ability (Kelly & Fiske, 1950, 1951; Deary, 2008; Roberts, Kuncel, Shiner, Caspi & Goldberg, 2007).
2. Level of education and jobs differ in their intellectual requirements (Gottfredson, 1997).
3. My colleagues and I have shown that there are also temperamental requirements for educational and job choice (Condon & Revelle, 2014; Revelle & Condon, 2012, 2015a; Revelle, Wilt & Condon, 2011; Wilt & Revelle, 2015)
4. We consider individual differences in Temperament, Ability, and Interests (TAI) as they relate to niche selection in choice of college major and in occupational choice (Bouchard, 1997; Hayes, 1962; Johnson, 2010). as well as to the second and third level of personality analysis (between individuals and between groups of individuals (Revelle & Condon, 2015b)).

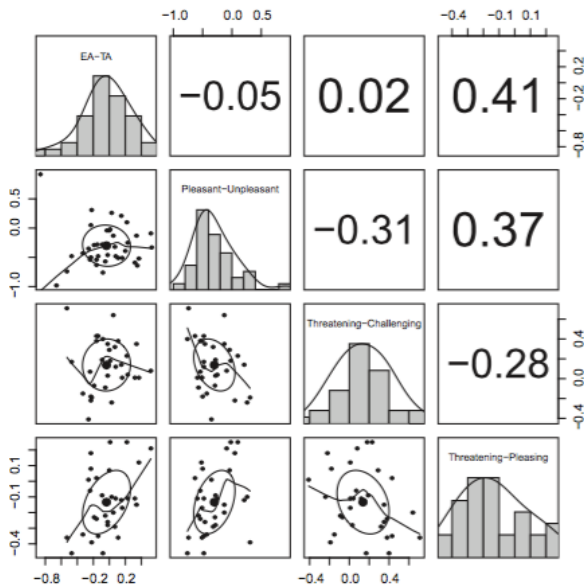
Measuring Affect, Behavior, Cognition and Desire within subjects

1. Personality is not just how people differ from each other, it is also how people differ in their Affect, Behavior, Cognition and Desire (ABCDs) over time and space.
2. This can be done using simple text messaging technology, or more delightfully elegant systems such as the Big EAR (Mehl, Pennebaker, Crow, Dabbs & Price, 2001; Mehl, Vazire, Ramirez-Esparza, Slatcher & Pennebaker, 2007; Mehl & Robbins, 2012) or Ryne Sherman's "Big Eye".
3. Most potential subjects carry around a very powerful data recorder: their cell phone.
4. Apps have been developed for smart phones to measure most anything
5. Even dumb phones can receive text messages.

Measuring the ABCDs with text messaging

1. In the past, if you wanted to measure the within person structure of affect, you needed to use paper and pencil diaries (Bolger, Davis & Rafaeli, 2003).
 - This was tedious for the participant
 - Difficult for the investigator.
2. Josh Wilt, Katie Funkhouser, Wiebke Bleidorn and I have used cell phone text messaging to measure affect, behavior, and meaning over time (Wilt, Funkhouser & Revelle, 2011; Wilt, Bleidorn & Revelle, 2016).
3. The basic observation is that the structure of affect within people is stable, is predicable, and differs across people. (see also (Rafaeli, Rogers & Revelle, 2007).
4. While for some people, Energetic Arousal is positively correlated with Tense Arousal, for others they are orthogonal, and for yet others they are negatively correlated. This correlation is, itself, predictable.

Structure of Affect within subjects varies by orientation to world



Measuring individual differences

1. A basic problem in the study of individual differences is that there are so many different constructs that interest us. These include constructs from at least four broad domains
 - Temperament
 - Ability
 - Interests
 - Character
2. Each domain has many constructs:
 - Dimensions of Temperament 2-3-5-6-15?
 - Structure of Ability ($g - g_f, g_c, V-P-R$)?
 - Hierarchical structure of interests people-things, RIASEC .
 - Range of possible measures of character.
3. But many important measures are proprietary.
4. In addition, showing the utility of TAIC measures requires criterion variables, and should include demographics.
5. Our solution: Use and/or develop open source temperament, ability, and interest items.

The International Personality Item Pool (Goldberg, 1999)

1. Perhaps one of the greatest contributions from Lew Goldberg has been his release of the International Personality Item Pool or IPIP (Goldberg, 1999) s <http://ipip.ori.org>.
2. The IPIP adapted a short stem item format developed in the doctoral dissertation of Hendriks (1997) and items from the Five Factor Personality Inventory developed in Groningen (Hendriks, Hofstee & De Raad, 1999).
3. Goldberg (1999) used about 750 items from the English version of the Groningen inventory, and has since supplemented them with many more new items in the same format.
4. The IPIP items have been translated into at least 39 languages by at least 65 different research teams. This includes Arabic, German, Farsi, Icelandic, Indonesian, Japanese, Korean, Mandarin, Polish, Portuguese, Russian, Serbian, Spanish, Turkish, Urdu, Slovenian and Swedish.

IPIP and other personality inventories

1. The IPIP was originally meant to be short stems to measure the Abridged Five Factor Circumplex structure of adjectives (Hofstee, de Raad & Goldberg, 1992) but also includes items targeted at most major personality tests.
2. Using a panel of roughly 1000 residents from Eugene-Springfield, Oregon, Goldberg administered his original IPIP items along with the NEO-PI-R (Costa & McCrae, 1992), the CPI (Gough & Bradley, 1996), the 16PF (Cattell & Stice, 1957), the MPQ (Tellegen & Waller, 2008), the Hogan PI (Hogan & Hogan, 1995), the TCI (Cloninger, Przybeck & Svrakic, 1994), the JPI-R (Jackson, 1983), and the 6FPQ (Jackson, Paunonen & Tremblay, 2000).
3. Goldberg then developed item stems that were highly correlated to the commercial inventories and put these into the public domain with the formation of the IPIP.
4. The items are available at <http://ipip.ori.org> and the Eugene-Springfield data are available from Goldberg.

What are the “Big 5”? Some representative items

Semantic analysis of many (although primarily European) languages suggest 5 broad factors of the ways in which we describe others.

Conscientiousness Complete my duties as soon as possible. Do things according to a plan. Like order.

Agreeableness Take advantage of others. (R) Am concerned about others. Sympathize with others' feelings.

Neuroticism Get upset easily. Get overwhelmed by emotions. Have frequent mood swings.

Openness/Intellect Am able to come up with new and different ideas. Am full of ideas. Have a rich vocabulary.

Extraversion Like mixing with people. Enjoy meeting new people. Am a talkative person. Am rather lively.

These are sometimes organized as the OCEAN of personality, alternatively, the CANOE of personality.

Extending the IPIP

1. In addition to the basic temperament items at the IPIP site, there are additional items to measure vocational interests (the ORVIS) (Pozzebon, Visser, Ashton, Lee & Goldberg, 2010) as well as avocational interests (Goldberg, 2010).
2. David Condon has expanded 2500 IPIP items to include the original IPIP items, the ORVIS, the ORAIS, as well as items from the EPQ (Eysenck, Eysenck & Barrett, 1985), the O*NET interest profile scales (Rounds, Su, Lewis & Rivkin, 2010). These, and other items make a total set of 4,300 items.
3. These are available at <https://sapa-project.org/MasterItemList/>.
4. Condon has also noted that although 18 different inventories (with 168 scales) have what appear to be 1,894 items, there are actually just 696 unique items. In addition, those “magic 696” cover between 57% to 85% of 10 additional inventories with 235 additional scales.

The International Cognitive Ability Resource

ICAR: Extending the IPIP to ability: IPIP:Personality::ICAR:Ability

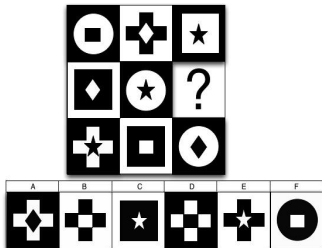
1. ICAR is an international collaboration to develop open source cognitive ability items.
2. Information at <http://www.icar-project.com/>
3. News letter at http://www.icar-project.com/ICAR_News_Issue_One.pdf
4. Key organizers who are coordinating the project:
 - Germany** Phillip Doebler (Münster and Ulm) and Heinz Holling (Münster)
 - U.K.** Luning Sun and John Rust (Cambridge)
 - U.S.A** William Revelle and David Condon (Northwestern)
5. Everyone is welcome to join this international collaboration.
6. Supported by Open Research Area (ORA) for the Social Sciences which includes participation from national funding agencies (Germany:DFG), (UK:ESRC), (US:NSF).

ICAR: Proof of concept

1. About 60 items were developed as part of a honors thesis at Northwestern by Melissa Liebert (Liebert, 2006) and meant to be “Google resistant” (answers are not available on the web).
 - This set was reported at conference in Krakow and in a subsequent book chapter (Revelle et al., 2010).
2. Subsequently David Condon developed some 3 Dimensional rotations items and did some extensive item analysis of the total set.
3. Condon & Revelle (2014) examined the first 60 publicly available items and validated them against self reported SAT exam scores as well as a small sample given the Shipley-2 (Shipley, 2009).
4. The original data set has been released to DataVerse (Condon & Revelle, 2015a) and has been published in the *Journal of Open Psychology Data* (Condon & Revelle, 2015c).
5. An example data set of 16 items with $N = 1,525$ is included as the *ability* dataset in the *psych* package.

Sample ICAR items

Matrix Reasoning



Verbal Reasoning

What number is one fifth of one fourth of one ninth of 900?

- (1) 2 (2) 3 (3) 4 (4) 5 (5) 6 (6) 7
- _____

If the day after tomorrow is two days before Thursday, then what day is it today?

- (1) Friday (2) Monday (3) Wednesday
(4) Saturday (5) Tuesday (6) Sunday

Letter and Number Series

In the following alphanumeric series, what letter comes next?

I J L O S

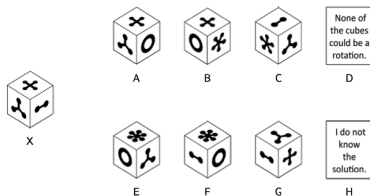
- (1) T (2) U (3) V (4) X (5) Y (6) Z
- _____

In the following alphanumeric series, what letter comes next?

Q S N P L

- (1) J (2) H (3) I (4) N (5) M (6) L

Three-Dimensional Rotation



Sample analysis of ICAR items

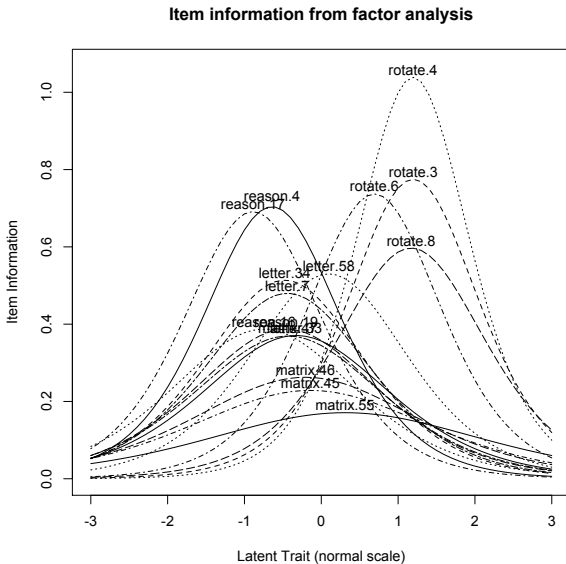
- Using basic R functions in the *psych* package (Revelle, 2015) we can evaluate the factor structure of the ICAR items.
- `irt.fa` will do a factor analysis of the items and report the statistics in terms of those statistics more commonly used in Item Response Theory.
 - The two parameters from factor analysis are item difficulty taken from the τ parameter from the tetrachoric correlation and the item factor loading λ of the matrix of tetrachoric correlations.

$$a = \frac{\lambda}{\sqrt{1 - \lambda^2}} \qquad \delta = \frac{\tau}{\sqrt{1 - \lambda^2}}$$

- The hierarchical structure of the ability items may be shown by factoring the factor intercorrelations.
- Loadings on a general factor may then be found by using the `omega` function which applies a Schmid Leiman transformation to the resulting higher level solution.

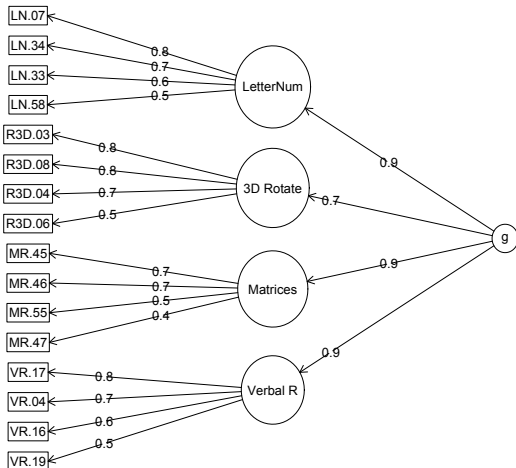


Item information curves for the 16 ICAR sample set



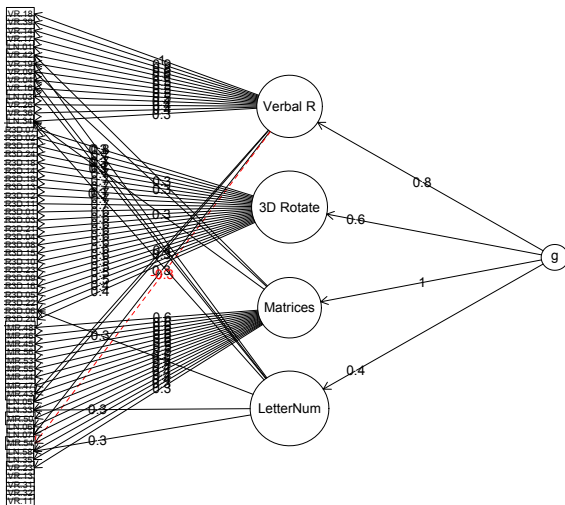
Structure of sample ICAR 16 items shows a clear 4 factor hierarchical solution $\omega_h = .87$

Omega Hierarchical for ICAR Sample Test



Structure of ICAR 60 items shows a messier 4 factor hierarchical solution $\omega_h = .76$

Hierarchical structure of ICAR60 items



Open materials

1. The International Personality Item Pool items (Goldberg, 1999) as well as the extended IPIP are in the public domain and are available to anyone for free.
2. The items from the International Cognitive Ability resource are also in the public domain and are available to registered users. (We are trying to keep the items relatively secure and do not put all of the actual items up on the web.)
 - We have a basic set of 60 ICAR items (Condon & Revelle, 2014) and the ICAR group is developing and validating item generators to automatically produce hundreds of each of a growing number of item types.
 - We encourage others to join us in this mission.
 - Go to <http://www.icar-project.com/> to register and for more information.

Part III: Open Methods

Method: Synthetic Aperture Personality Assessment (SAPA)

Measuring individual differences: the tradeoff between breadth versus depth

Synthetic Aperture Astronomy

SAPA theory

Sample items as well as people

Covariance algebra

SAPA: practice

Open source software comes to the rescue

Part IV: Open Data

Four types of openness:

1. Open source software: The R project (R Core Team, 2015)
2. Open source materials:
 - The International Personality Item Pool (IPIP) (Goldberg, 1999)
 - The International Cognitive Ability Resource (ICAR) (Condon & Revelle, 2014)
3. Open source methodology: The Synthetic Aperture Personality Assessment Project (Revelle et al., 2010, 2016)
4. Open source data:
 - Data from the ICAR project (Condon & Revelle, 2016, 2015a)
 - Data from SAPA studies (Condon & Revelle, 2015c,b)

In the process of summarizing the last several years of research, I will show how we use open source software, items, and methods and then share them with the world.

Measuring individual differences

1. A basic problem in the study of individual differences is that there are so many different constructs that interest us. These include constructs from at least four broad domains
 - Temperament
 - Ability
 - Interests
 - Character
2. Each domain has many constructs
 - Dimensions of Temperament 2-3-5-6-?
 - Structure of Ability: $g - g_f, g_c, V-P-R$?
 - Hierarchical structure of interests people-things RIASEC
 - Range of possible measures of character
3. In addition, showing the utility of TAIC measures requires criterion variables

Breadth vs. depth of measurement

1. Factor structure of domains needs multiple constructs to define structure.
2. Each construct needs multiple items to measure reliably.
3. This leads to an explosion of potential items .
4. But, people are willing to only answer a limited number of items.
5. This leads to the use of short and shorter forms (the NEO-PI-R with 300, the IPIP big 5 with 100, the BFI with 44 items, the TIPI with 10) to include as part of other surveys.

Many items versus many people

1. Not only do we want many items, we also want many people.
2. Resolution (fidelity) goes up with sample size, N (standard errors are a function of \sqrt{N})

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N-1}} \quad \sigma_r = \frac{1-r^2}{\sqrt{N-2}}$$

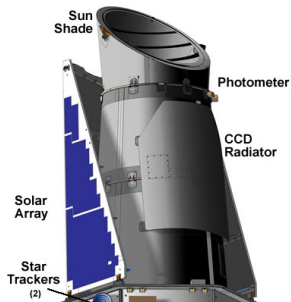
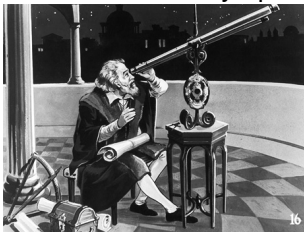
3. Also increases as number of items, n, measuring each construct (reliability as well as signal/noise ratio varies as number of items and average correlation of the items)

$$\lambda_3 = \alpha = \frac{n\bar{r}}{1 + (n-1)\bar{r}} \quad s/n = \frac{n\bar{r}}{(1-n\bar{r})}$$

4. Thus, we need to increase N as well as n. But how?

A short diversion: the history of optical telescopes

Resolution varies by aperture diameter (bigger is better)



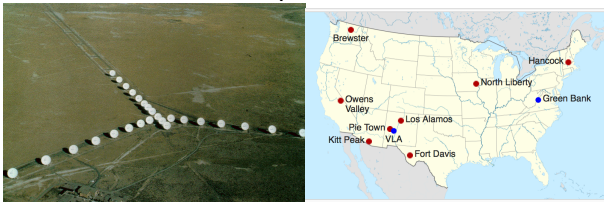


A short diversion: history of radio telescopes

Resolution varies by aperture diameter (bigger is still better)



Aperture can be synthetically increased across multiple telescopes
or even multiple observatories



Can we increase N and n at the same time?

1. Frederic Lord (1955) introduced the concept of sampling people as well as items.
2. Apply basic sampling theory to include not just people (well known) but also to sample items within a domain (less well known).
3. Basic principle of Item Response Theory and tailored tests.
4. Used by Educational Testing Service (ETS) to pilot items.
5. Used by Programme for International Student Assessment (PISA) in incomplete block design (Anderson, Lin, Treagust, Ross & Yore, 2007).
6. Can we use this procedure for the study of individual differences without being a large company?
7. Yes, apply the techniques of radio astronomy to combine measures synthetically and take advantage of the web.

Subjects are expensive, so are items

1. In a survey such as Amazon's Mechanical Turk (MTURK), we need to pay by the person and by the item.
2. Why give each person the same items? Sample items, as we sample people.
3. Synthetically combine data across subjects and across items. This will imply a missing data structure which is
 - Missing Completely At Random (MCAR), or even more descriptively:
 - Massively Missing Completely at Random (MMCAR)
4. This is the essence of Synthetic Aperture Personality Assessment (SAPA).



3 Methods of collecting 256 subject * items data

a) 8 x 32 complete

46213634521143453443645331212414
 21243623166421516154432261516513
 51661351155165463622224435623344
 1141343362332215612152135614522
 25353121264561433433232246526411
 61335154566424114612641225353516
 24634342151536242425413513435116
 11554654453123111162423325516334

b) 32 x 8 complete

46323114
 25443314
 43315423
 26314145
 41435614
 42236153
 62421344
 35234443
 34514166
 63415154
 44441342
 13514321
 66365663
 12264546
 31466135
 32645514
 66151251
 14411441
 62443636
 33316236
 63325425
 11531126
 61155546
 33245361
 52241654
 63212356
 24414663
 63661414
 45555223
 14364433
 21461416
 33232365

c) 32 x 32 MCAR $p=.25$

. . 3 . . 2 . . 6 4 . 55 44
 4 . 6 . 45 . 3 . 4 . 6 1
 6 . 3 6 . 1 6 . 2 5 . 6
 3522 5 . 3 3 5
 3 . 2 . 2 3 . 2 65 . 5 .
 51 324 23 5
 552 25 . . 54 . 5
 . . 44 . 4 . 5 3 . 6 6 3
 61 . 523 . 2 2 3
 5 42 . 4 . 6 . 5 61 .
 3 3 . 6 . 1 . 4 . . 1 . 5 5 .
 1 54 2 . 4 . 33 . 6
 4 52 . 6 44 . 3 2
 . 44 . . 1 1 . 42 5 . 1
 . 1 . 3 2 . 3 . 521 6
 3 . 142 22 12 .
 . 4 . . 2 3 . 162 . . 4 4
 . 4 . 6 . 3 . 4 . . 1 5 . 33
 5 243 . 5 41 1 . .
 . 5 . 3 . 4 . . 4 . 4 . 5 . 1 4 .
 4 3 . 5 . 2 64 . 4 . 4 .
 . . 1 . 1 . 2 . . 6 4 55 . . 2 . .
 3 . 2 . . 53 2 . 2 . 3 . 3
 1 . . 2 . 43 . . 3 . 13 5 .
 . . 2 4 . 54 . . 2 . 3 . 62
 22 332 . 1 5 6
 . . 5 . 3 . 4 3 5 . 241
 63 . 1 6 . . 5 . 4 . 2 . . 5
 . 2 . 4 . 5 52 . 4 44
 2 . 55 2 6 6 55
 . 5 4 6341 . 4 . 2
 55 5 45 3 . 32

Synthetic Aperture Personality Assessment

1. Give each participant a random sample of pn items taken from a larger pool of n items.
2. Find covariances based upon “pairwise complete data”.
3. Find scales based upon basic covariance algebra.
 - Let the raw data be the matrix \mathbf{X} with N observations converted to deviation scores.
 - Then the item variance covariance matrix is $\mathbf{C} = \mathbf{X}\mathbf{X}'N^{-1}$
 - and scale scores, \mathbf{S} are found by $\mathbf{S} = \mathbf{K}'\mathbf{X}$.
 - \mathbf{K} is a keying matrix, with $\mathbf{K}_{ij} = 1$ if *item*_{*i*} is to be scored in the positive direction for scale *j*, 0 if it is not to be scored, and -1 if it is to be scored in the negative direction.
 - In this case, the covariance between scales, \mathbf{C}_s , is

$$\mathbf{C}_s = \mathbf{K}'\mathbf{X}(\mathbf{K}'\mathbf{X})'N^{-1} = \mathbf{K}'\mathbf{X}\mathbf{X}'\mathbf{K}N^{-1} = \mathbf{K}'\mathbf{C}\mathbf{K}. \quad (1)$$

4. That is, we can find the correlations/covariances between scales from the item covariances, not the raw items.

SAPA standard errors and effective sample size

1. When forming synthetic scales from MMCAR based items, the standard error of correlations decreases as a function of the Total number of subjects (N), the percentage of items samples (p), *and* the number of items forming the scale (n).
2. Ashley Brown has shown this quite clearly by simulation (Brown, 2014).
3. A good way to visualize this is to examine the standard error of correlations as a function of N, p, and n.
4. An even more dramatic way is to plot the *Effective Sample Size* (N_{eff}) which because

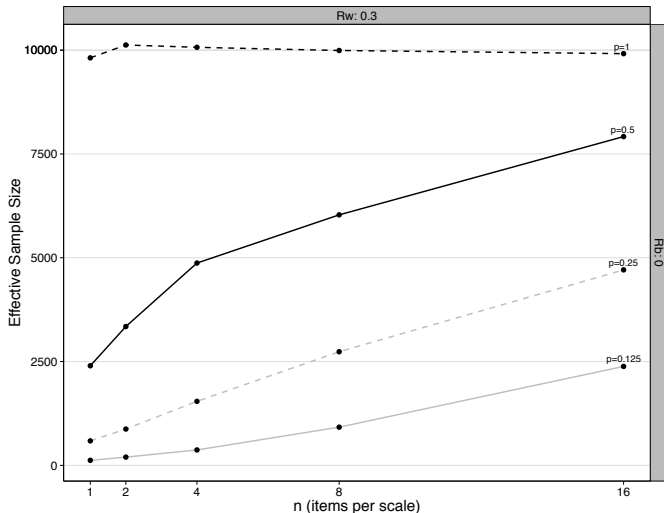
$$\sigma_r = \frac{1 - r^2}{\sqrt{N - 2}} \text{ is merely } N_{eff} = \frac{(1 - r^2)^2}{\sigma_r^2} + 2$$



Effective sample size varies by the size of the composite scale.

Simulating $N = 10,000$ with probability of any item (Brown, 2014)

($p = .125, .25, .5, \text{ or } 1$) and items in the composite 1, 2, 4, 8, 16.



SAPA is not magic: We can obtain high accuracy at the structure level but accuracy is much lower at the single subject level

1. Reliability of composite scales is high when formed from synthetic matrices $\mathbf{C}_s = \mathbf{K}'\mathbf{C}\mathbf{K}$ because the number of items per scale/per subject is the nominal amount.
2. Reliability of single scores is much less because very few items measuring a single trait are given to a single subject $\mathbf{S} = \mathbf{K}'\mathbf{X}$.
3. However, the precision of the estimate of subject means (\bar{x}) is high because $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{Np-1}}$ and Np is large.
4. SAPA technique is very powerful for research of structure, but less powerful for research based upon single subjects.
5. Particularly useful in web based surveys with many subjects when we are limited in the number of items we can administer and where we are trying to increase our domain validity.

How does it work?

1. Give our basic belief in open science, we use public domain items, open source software:
 - Apache webserver, MySQL data bases, PHP and HTML5 web tools, R for statistics.
 - Extensive coding in PHP and MySQL to present item sets in random fashion (Joshua Wilt, David Condon, Jason French)
 - Code written for psychometric measurement and scale construction as implemented in the *psych* package (Revelle, 2015) using R (R Core Team, 2015)
2. Domains measured and item sources
 - Temperament items taken from International Personality Item Pool (IPIP) (Goldberg, 1999) (ipip.ori.org) and supplemented with other items.
 - Ability items have been validated (Condon & Revelle, 2014) as part of the International Cognitive Ability Resource Project (ICAR-project.org). (ICAR:Ability::IPIP:Temperament)
 - Interest items taken from Oregon Vocational Interest Survey (ORVIS) (Pozzebon et al., 2010)

SAPA overview

1. A “Personality Test” is included as a resource at the <http://personality-project.org> and gives feedback to all participants.
2. Some participants then link their feedback to their social media sites which then appeals to yet more to take it.
3. Some professors assign it to their students in various classes.
4. About 120-180 people per day from around the world visit the personality-project.org or sapa-project.org websites. This does not sound like much, but over a year, we get around 40,000-50,000 participants.
5. This is much less than the $6 * 10^6$ subjects that Gosling and others are reporting, but we are talking about structure based upon 700-1000 items instead of a limited 44 BFI items.
6. This the benefit of random sampling.



The SAPA Project

Take the test.

Explore your personality.

Advance the study of individual differences.

[Start the test](#)

[More info](#)



FAQ about the test

Is it long? (not really) Is it free? (yes)



The research behind SAPA

How was the test developed?



Individual Differences

Learn more about differential psychology.

How does it work?: part II

1. Participants find us by searching web for “personality tests”, etc. and find `personality-project.org` or `sapa-project.org`

2. Each participant is given a number of web pages

Consent Form Basic description of project and question whether they have taken test before.

Demographics Age, sex, height, weight, education, parental education, country, state, ZipCode (if US), ...

TAIC questions Temperament/Ability/Interest questions (25 per page, 21 T/I, 4 Ability per page)

Continuation pages After each page, told that feedback will be more accurate if they keep going.

Optional modules Creativity, Peer ratings, interests, ...

Feedback Personality feedback based upon scores on temperament items.

3. Results are stored (page by page) on the MySQL server.

How does it work: part III

1. Various data cleaning scripts run using the *SAPA-tools* package (French & Condon, 2015) in R.
 - Screen for duplicate responses based upon a Random Identification Number issued when subjects start the page. We drop all subsequent pages.
 - Screen for subjects < 14 or > 90 .
2. Subsequent analyses are done primarily using functions in the *psych* package (Revelle, 2015) for R.
3. Analyses are done at multiple levels:
 - 3.1 At the item and scale covariance level to examine the structure of items
 - 3.2 At the multiple levels of aggregation: zip code, state, college major, occupation. This requires finding individual level scores and then examining the structure of group means through basic multi-level techniques.



Part IV: Open Data

Part IV: Open Data

Demographics of the SAPA data

- Sample characteristics

- Web data are not random samples

Personality Structure

- Are the “Big Five” really big?

- Temperament, Ability and Interests,

Temperament, Ability and Interests: Occupational Choice

- Pooled correlations \neq within group or between group correlations

- Occupational Choice as niche selection

Summary: Open Science



Four types of openness:

1. Open source software: The R project (R Core Team, 2015)
2. Open source materials:
 - The International Personality Item Pool (IPIP) (Goldberg, 1999)
 - The International Cognitive Ability Resource (ICAR) (Condon & Revelle, 2014)
3. Open source methodology: The Synthetic Aperture Personality Assessment Project (Revelle et al., 2010, 2016)
4. Open source data:
 - Data from the ICAR project (Condon & Revelle, 2016, 2015a)
 - Data from SAPA studies (Condon & Revelle, 2015c,b)

In the process of summarizing the last several years of research, I will show how we use open source software, items, and methods and then share them with the world.



Our data

Time Frame Data collected at personality-project.org and sapa-project.org from August 18, 2010 to December 22, 2015

Subjects N = 207,002 (77,550 males, 129,451 females)

Materials 953 items (696 temperament, 60 ability, 152 interests, 45 demographic)

Scales used 15 Temperament, 4 Ability, 6 Interests

N in workforce N = 80,317

N students N = 86,348

Occupations 973 separate occupations, following a Pareto distribution with $\approx 80\%$ represented by the top 20% of occupations

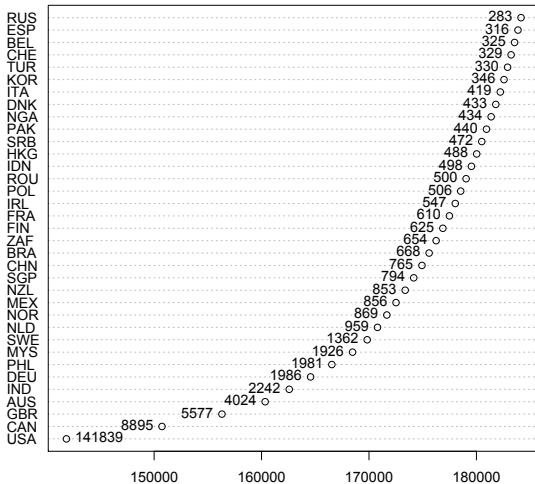
$N \geq 75$ 195 occupations for 55,902 participants

$N \geq 100$ 116 college majors for 130,584 participants



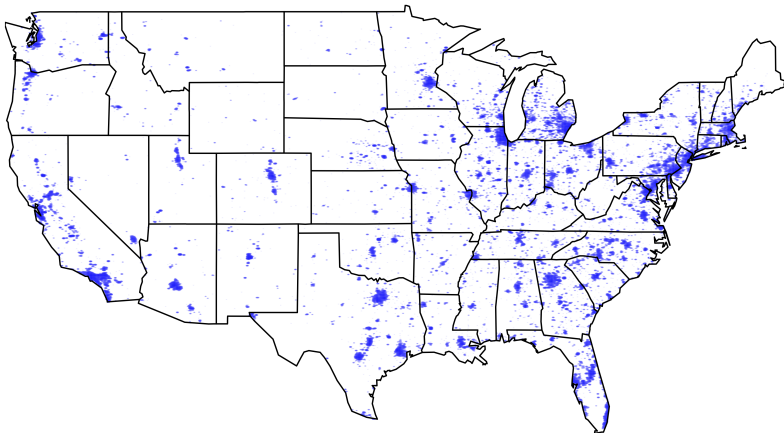
The top 35 countries account for 89% of the sample

The top 35 countries account for 89% of the sample





Where do they come from? US SAPA data by zipcodes



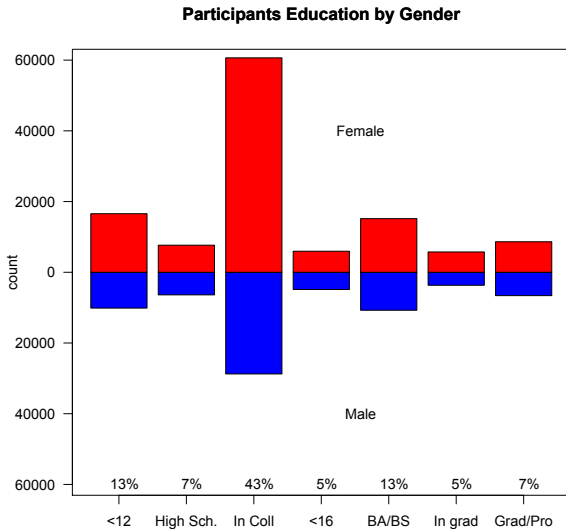


US lights (from NASA)



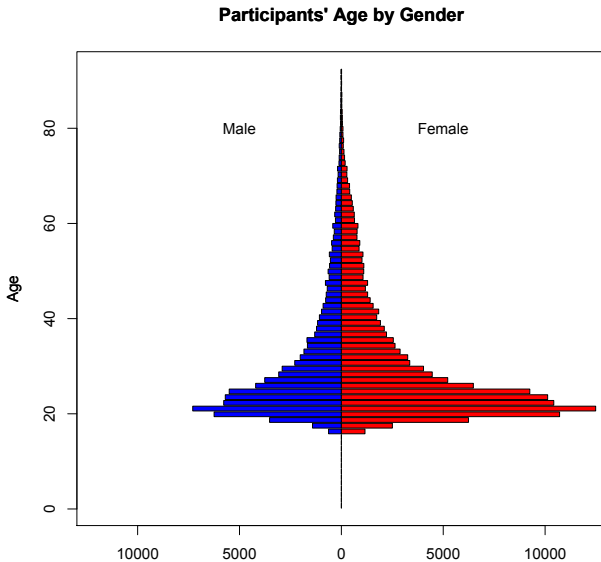


Not a random sample of either education or gender (62% female)





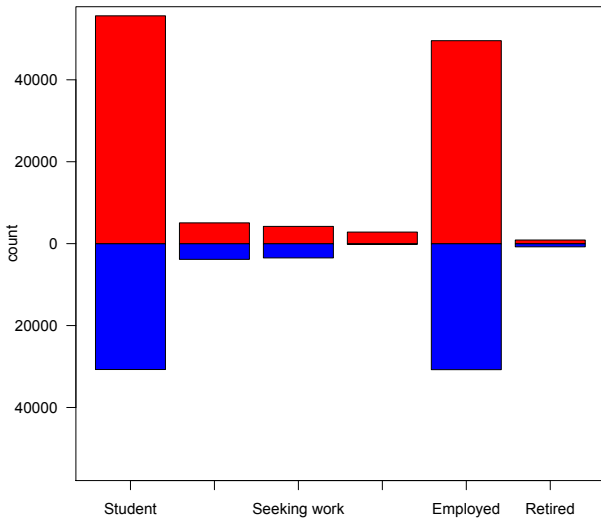
Not a random sample of either age or gender (62% female)





86,348 students, 80,317 employed)

Employment by Gender





Are the “Big Five” really big?

1. There is a “consensus” about the proper number of factors/components of personality (Goldberg, 1990, 1992; Hofstee et al., 1992).
2. This seems to match life challenges of Getting Along and Getting Ahead
 - Conscientiousness Work
 - Agreeableness Love
 - Neuroticism Effective functioning in many domains
 - Openness/Intellect Play
 - Extraversion Leadership
3. Additional work has been done on the same 800-1000 person Eugene-Springfield sample and suggests a hierarchical structure (e.g., the BFAS) (DeYoung, Quilty & Peterson, 2007).
4. But what happens with a larger and more diverse sample?



Multiple solutions to the dimensionality of temperament

- 2) Digman “alpha and beta” (Digman, 1997), DeYoung “stability and plasticity” (DeYoung, Peterson & Higgins, 2002)
 - 3) Eysenck “Giant 3” (Eysenck, 1994) PEN
 - 5) The “Big 5” (Digman, 1990; Goldberg, 1990) CANOE/OCEAN/ I ... V (Norman, 1963; Tupes & Christal, 1961)
 - 6) The HEXACO 6 (Lee & Ashton, 2004; Ashton, Lee & Goldberg, 2007)
 - 7-9) Tellegen 7-9 (Tellegen & Waller, 2008)
 - 8-9) Comrey 8-9 (Comrey, 2008)
 - 16) Cattell 16 Personality Factors (Cattell, 1957)
1. Condon (2014, 2015) examined 696 non-overlapping items from IPIP:100, IPIP:NEO, IPIP:MSQ, BFAS, EPQ, etc. (Goldberg, 1999; DeYoung et al., 2007; Eysenck et al., 1985)
Found meaningful 3, 5, 15, (and now) 27 factor solutions.
 2. The Condon 3/5/15/27 form a heterarchical and non hierarchical structure (i.e., lower levels are not cleanly nested in higher levels.)



Personality shows a heterarchical even fractal structure

1. David Condon (2014, 2015) and in prep has shown:
2. The structure of 696 personality items given to 100-200,000 participants does not show a clean organization.
3. The number of factors problems (and its non-solutions) will break the heart of most investigators.
4. No clear structure at any level.
5. Equally compelling fits at 3/5/15/27 factors.
6. The larger the sample size, the greater the resolution and there is always something more there.
7. Articles showing “incremental validity beyond the Big Five” are just as likely to show incremental validity beyond any number of factors.
8. We think of this as showing a fractal structure: equal level of complexity at any level of resolution.



The best items from the 15 scale solution

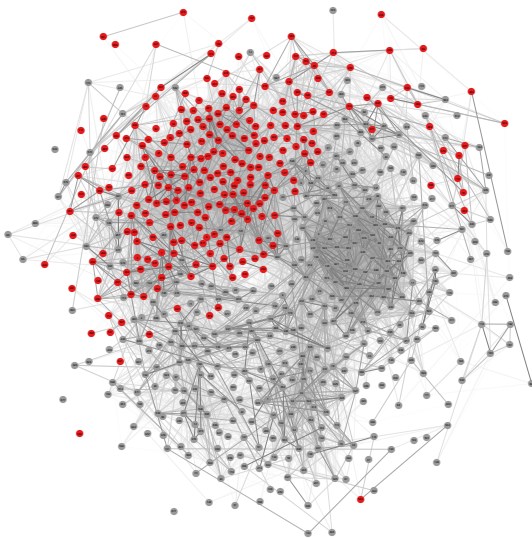
Table : Sample items from the Short Personality Inventory 15 factor solution

Each scale has 8-10 items

SPI	Item	Item
Fear	Panic easily.	Begin to panic when there is danger.
Volatility	Get irritated easily.	Lose my temper.
Outlook	Dislike myself.	Feel a sense of worthlessness or hopelessness.
Compassion	Sympathize with others feelings.	Am sensitive to the needs of others.
Trust	Trust others	Trust what people say.
Easygoing	Let things proceed at their own pace.	Take things as they come.
Industrious	Start tasks right away.	Get chores done right away.
Mach	Use others for my own ends.	Cheat to get ahead.
Impulsivity	Act without thinking.	Do things without thinking of the consequences.
Sociability	Am mostly quiet when with other people.	Tend to keep in the background on social occasions.
Boldness	Love dangerous situations.	Take risks.
Serious	Seldom joke around.	Am not easily amused.
Conventional	Don't like the idea of change.	Prefer to stick with things that I know.
Intellectual	Am quick to understand things.	Catch on to things quickly.
Open	Enjoy thinking about things.	Love to reflect on things.

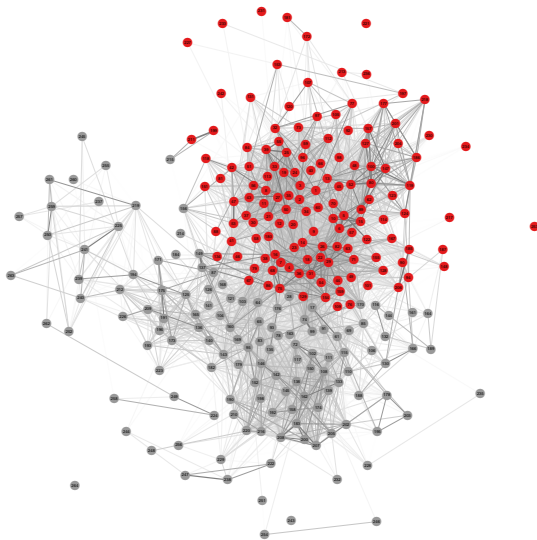


The correlations of 696 personality items



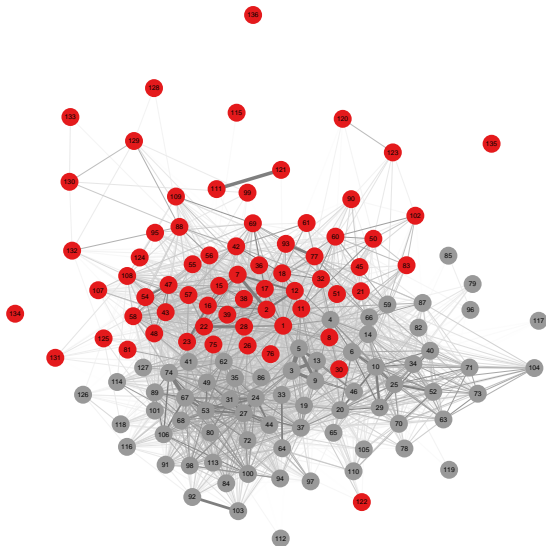


Factoring the items on the first factor of the 696



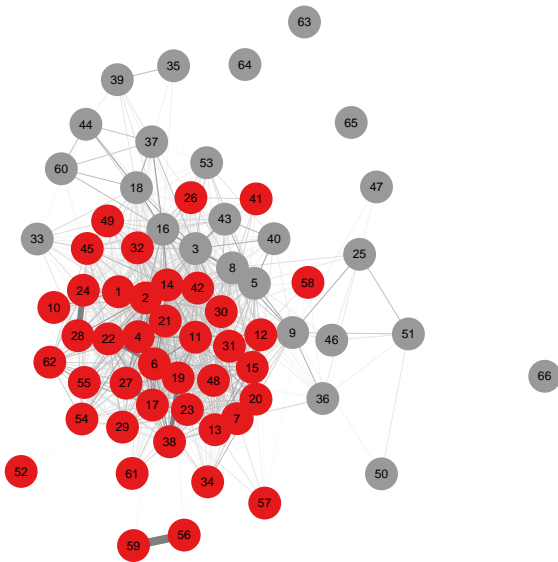


And doing it again





And again





6 factors of interests

1. 6 factors from the O*NET interest profiler scales (60 items; Rounds et al., 2010)
2. 8 factor Oregon Vocational Interest Scales (92 items; Pozzebon et al., 2010)
3. Oregon Avocational Interest Scales (199 items; Goldberg, 2010)
4. Formed into 6 scales fitting a “RIASEC” structure (60 items)
 - Realistic** “Like to work with tools and machinery.”
 - Investigative** “Would like to do laboratory tests to identify diseases.”
 - Artistic** “Would like to write short stories or novels.”
 - Social** “Would like to help conduct a group therapy session.”
 - Enterprising** “Would like to be the chief executive of a large company.”
 - Clerical** “ Would like to keep inventory records”



The correlational structure of scales are found from the SAPA item correlations

1. Given the raw data matrix, we can find the covariances (using pairwise complete data) and then find the scale intercorrelations.
2. The correlations with scales with overlapping items can be corrected for overlap using a correction by Bashaw & Anderson Jr (1967); Cureton (1966)
3. We can do this for 696 Temperament items, 60 Ability items, and 152 Interest items.
4. Although we have 150 different scales, I am reporting just IPIP based Big 5, 4 ICAR ability scales, and the 6 RIASEC scales.
5. We can also drill down to the “Aspects Level” (DeYoung et al., 2007), the “facet level” (Costa, McCrae & Dye, 1991), the “nuances” (Mottus, 2016; Ottus, Kandler, Bleidorn, Riemann & McCrae, 2016) or items.



Personality at 3 levels of analysis (Revelle & Condon, 2015b)

Personality can be examined at three levels of analysis

1. Personality as a unique temporal signature of one's Affect, Behavior, Cognition and Desires (ABCDs) as they change over time and space within a single individual.
 - Measuring within person patterning requires repeated measures on single subjects over time. We do this with open source text messaging procedures e.g., (Wilt et al., 2011; Wilt, 2014).
2. Personality is also how people differ in their patterning of the ABCDs between people.
 - This can be multilevel modeling of data collected within subjects showing that the correlational structure within subjects differs across subjects (Wilt et al., 2011; Revelle & Wilt, 2015).
 - It is also the more conventional structure of personality items as collected from the SAPA project.
3. But people choose groups such as college major or occupation based upon their unique aptitudes and appetites.
 - We can analyze this niche selection in terms of the covariance of the mean personality of the group.



TAI for groups is not the same as TAI for individuals

- How do occupational groups or college majors differ on TAI?
 - The mean scores for groups allow us to compare the groups
 - But it is the structure of these group means that are particularly interesting for they allow us to examine niche selection.
- Overall correlation is a function of within group correlations and between group correlations.
- Correlations of aggregate scores $r_{xy_{bg}}$ (between groups) \neq aggregate of correlations $r_{xy_{wg}}$ (within groups)
- The overall correlation r_{xy} is a function of the within and the between correlations

$$r_{xy} = \eta_{x_{wg}} * \eta_{y_{wg}} * r_{xy_{wg}} + \eta_{x_{bg}} * \eta_{y_{bg}} * r_{xy_{bg}}$$

- These multi level correlations sometimes lead to what is known as the Yule-Simpson paradox (Kievit, Frankenhuis, Waldorp & Borsboom,

2013; Simpson, 1951; Yule, 1903)

- These are independent and useful information.



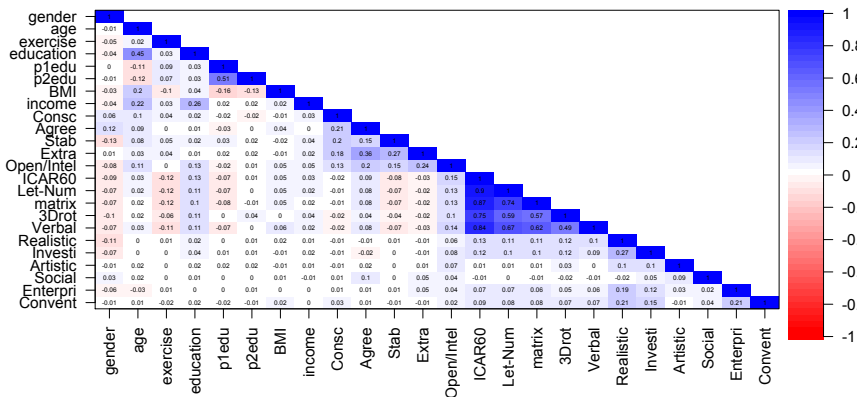
Temperament, Ability, and Interests – within and between groups

1. Examined the factor structure of the TAI scales at the normal, between subjects (across groups) level.
 - This produces the normal factor structure of temperament, of ability and of interests
 - Can show these correlations as a “heatmap”
2. But when analyzing the structure of the mean scores for each of 148 majors or 196 occupational groups (minimum size of 75 members), the structure is drastically different.
 - Several dimensions of temperament and interests are now negatively correlated with ability, others are orthogonal
 - Can also show these correlations as a “heatmap”.
 - In that some correlations change sign, this is not an artifact of aggregation.



Subject Level data of 8 criteria, 5 personality scales, 4 ability, 6 interests,

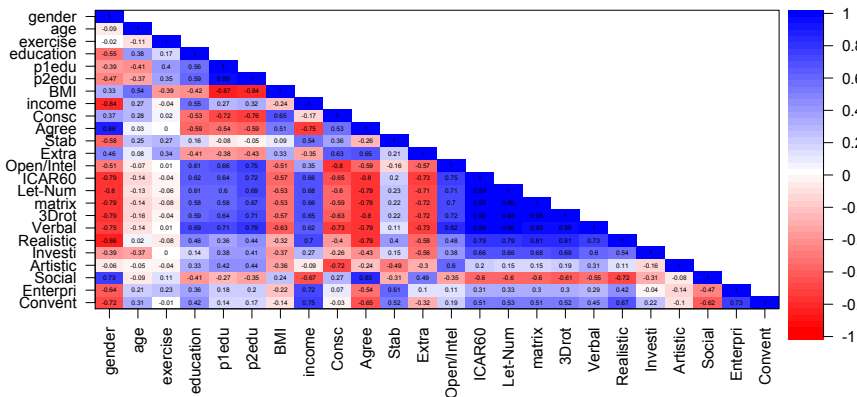
Within group correlations of Criteria, Temperament, Ability, Interests





Group Level data of 8 criteria, 5 personality scales, 4 ability and 6 interests”

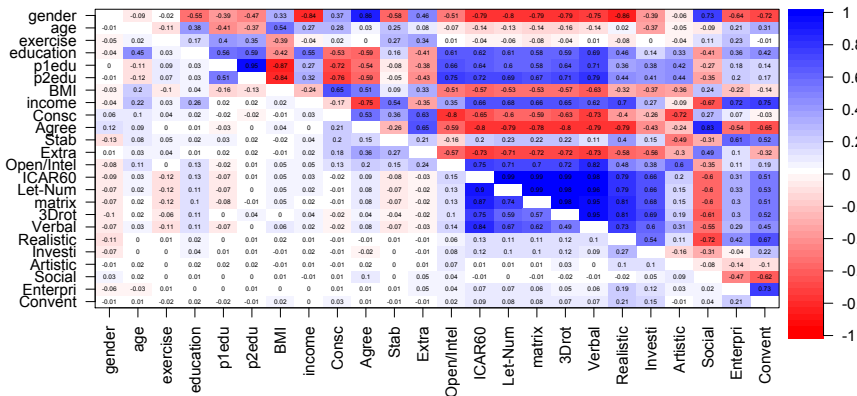
Between group correlations of Criteria, Temperament, Ability, Interests





Compare the two correlation matrices—college majors 8 criteria, 5 personality scales, 4 ability and 6 interests

Within group (below diagonal), Between groups (above diagonal)



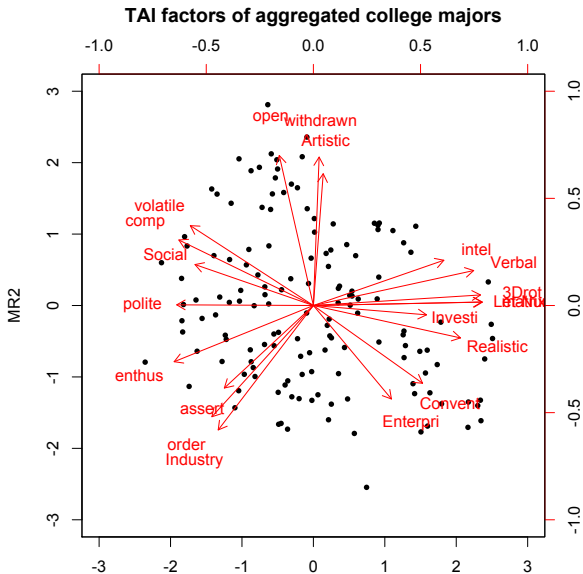


Niche selection

1. College majors and occupations differ systematically in the intellectual Ability they require.
2. But they also differ in the Interests and Temperament they require.
3. Examine the structure of the mean personality scale scores for college majors and for occupations.
4. The space is 2-3 dimensional.
5. A simple two factor solution shows that high ability can trade off for low Industry or Conscientiousness and that Boldness (low Anxiety) and Realistic interests differs from high Anxiety and Social interests.
6. We can examine the extent these two dimensions relate to a variety of criteria using factor extension (Dwyer, 1937; Horn, 1973)



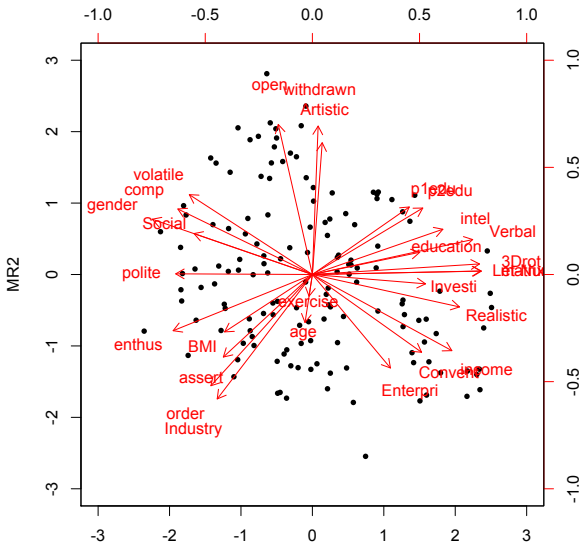
Biplot of a two factor solution to the group level major data





Add 8 criteria to the extended factor solution of the group data

TAI factors of aggregated majors extended to criteria variables



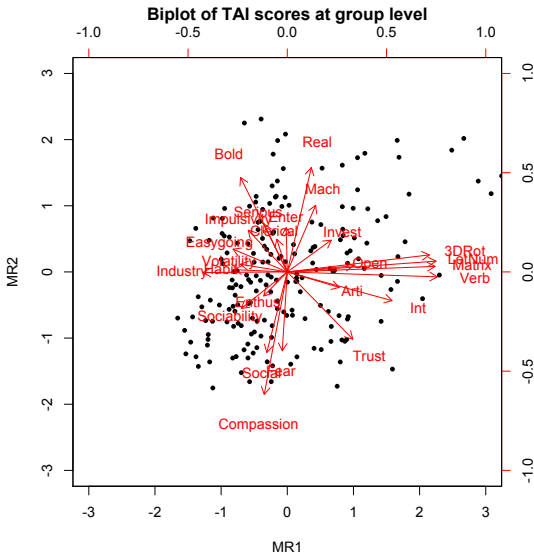


Occupational choice

1. Just as people assort themselves into college majors based upon the appetites and their aptitudes, so they assort themselves into occupations.
2. Some occupations require getting along, some emphasize getting ahead, some require both
3. I show a 2 dimensional solution and then using factor extension, add in gender to the model
4. The final figure is not meant to be read, but rather seen as the range of occupations covered.

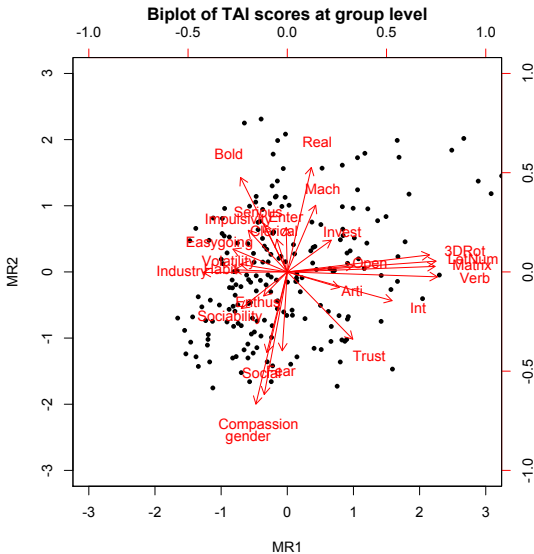


Biplot of a two factor solution to the group level data



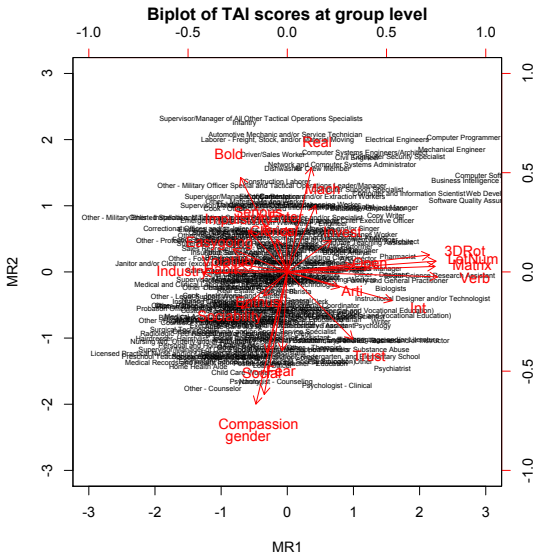


Biplot of a two factor solution to the group level data





Biplot of a two factor solution to the group level data



Part VI

Open Scientific Research

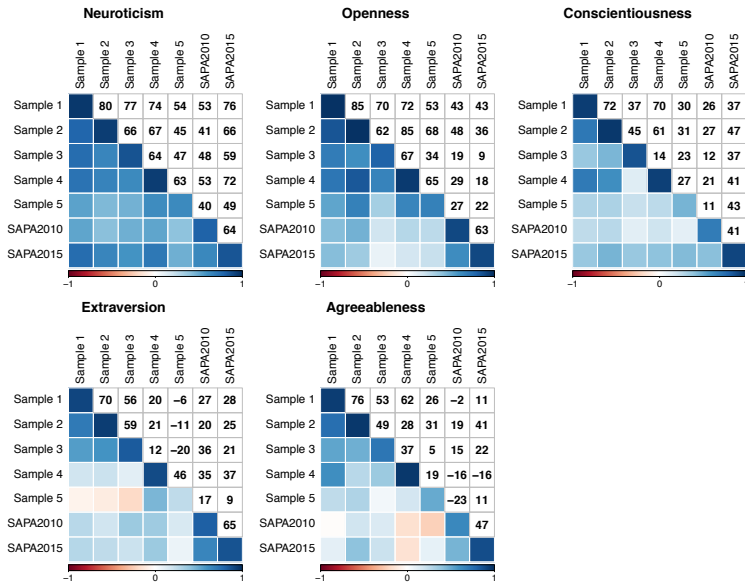
Replicability of personality research

1. To what extent is personality research replicable?
2. By using open methods and open data, we allow ourselves to test for replicability

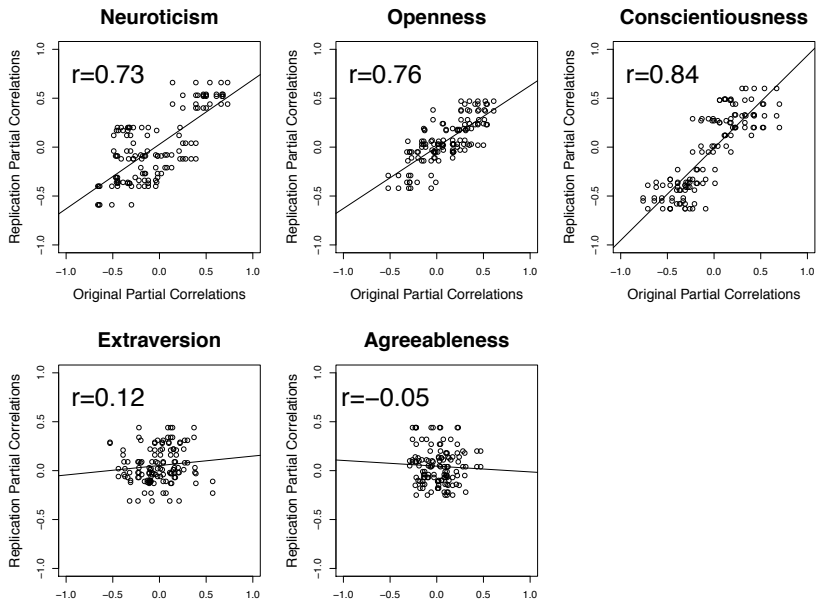
State level personality differences: Are they replicable?

1. Rentfrow and Gosling have reported data from several large internet surveys (much larger than ours).
2. Are they replicable?
3. With data shared from Gosling and Rentfrow, we were able to examine how much the 5 studies they report match our findings.
4. State differences are small, but reliable for some measures, not all.
5. We have looked at the basic state differences in personality mean scores
6. Also examined 18 different behavioral criterion (e.g., crime rates, state income, state well being, state liberalism, etc.)

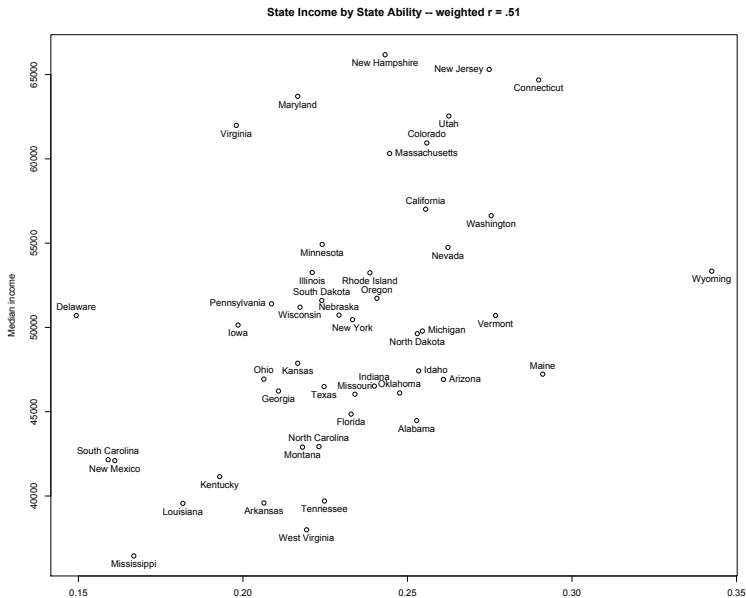
Replicability of state differences depends upon the trait



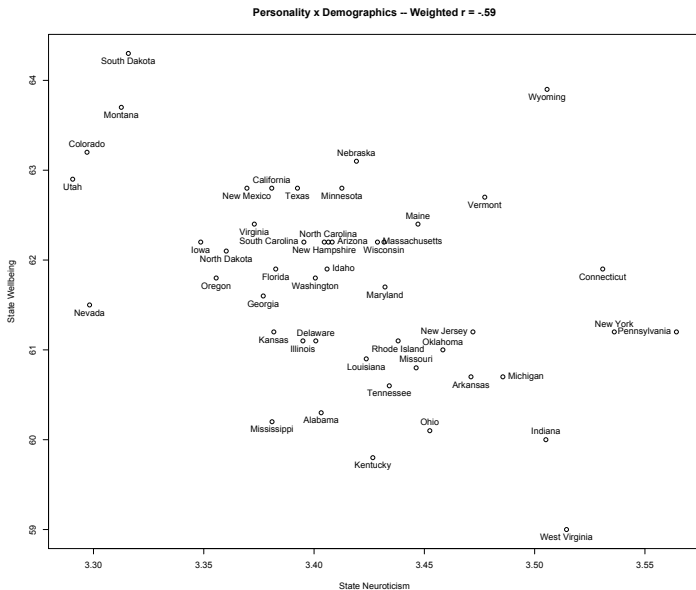
Replicability of personality by state demographics depends upon trait



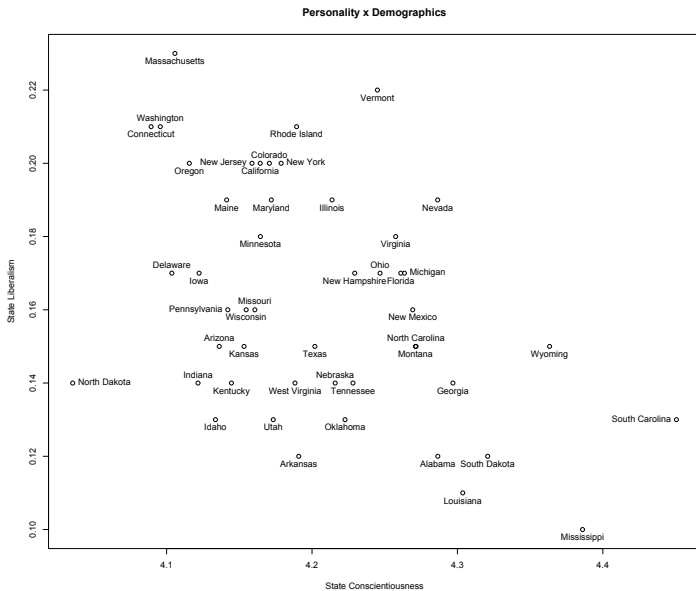
State level: Income varies by iq (weighted r = .51)



State level: Well being varies by neuroticism (weighted r = -.59)



State level: Liberalism varies by Conscientiousness ($r = -.42$)



Summary and Conclusions

1. Ability, temperament and interests all provide useful information about human personality.
2. Intellectual and personality development is the process of experiencing and choosing niches.
3. When we describe the intellectual requirements of a profession or a college major, we should not ignore that appropriate interests and temperaments guide occupational choice.
4. We need to consider appetites along with aptitudes.
5. The statistics, materials, methods, and data from all of these studies are done using Open Source Science.
6. Join us in this journey.
7. For more information and for these slides go to <http://personality-project.org/sapa.html>

- Anderson, C. J., Bahník, Š., Barnett-Cowan, M., Bosco, F. A., Chandler, J., Chartier, C. R., & Cheung, F. (2016). Response to comment on “estimating the reproducibility of psychological science”. *Science*, *351*(6277), 1037–1037.
- Anderson, J., Lin, H., Treagust, D., Ross, S., & Yore, L. (2007). Using large-scale assessment datasets for research in science and mathematics education: Programme for International Student Assessment (PISA). *International Journal of Science and Mathematics Education*, *5*(4), 591–614.
- Ashton, M. C., Lee, K., & Goldberg, L. R. (2007). The IPIP-HEXACO scales: An alternative, public-domain measure of the personality constructs in the HEXACO model. *Personality and Individual Differences*, *42*(8), 1515–1526.
- Bashaw, W. & Anderson Jr, H. E. (1967). A correction for replicated error in correlation coefficients. *Psychometrika*, *32*(4), 435–441.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods:

Capturing life as it is lived. *Annual Review of Psychology*, 54, 579–616.

Bouchard, T. (1997). Experience producing drive theory: how genes drive experience and shape personality. *Acta Paediatrica*, 86(S422), 60–64.

Brown, A. D. (2014). Simulating the MMCAR method: An examination of precision and bias in synthetic correlations when data are ‘massively missing completely at random’. Master’s thesis, Northwestern University, Evanston, Illinois.

Cattell, R. B. (1957). *Personality and motivation structure and measurement*. Oxford, England: World Book Co.

Cattell, R. B. & Stice, G. (1957). *Handbook for the Sixteen Personality Factor Questionnaire*. Champaign, Ill.: Institute for Ability and Personality Testing.

Cloninger, C. R., Przybeck, T. R., & Svrakic, D. M. (1994). *The Temperament and Character Inventory (TCI): A guide to its*

development and use. center for psychobiology of personality, Washington University St. Louis, MO.

- Comrey, A. L. (2008). The Comrey Personality Scales. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *Sage handbook of personality theory and testing: Personality measurement and assessment*, volume II (pp. 113–134). London: Sage.
- Condon, D. M. (2014). *An organizational framework for the psychological individual differences: Integrating the affective, cognitive, and conative domains*. PhD thesis, Northwestern University.
- Condon, D. M. (2015). The many little items of “big five” measures: Hierarchy, heterarchy, and predictive utility in personality structure. Long Beach, California. Society of Personality and Social Psychology.
- Condon, D. M. & Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence*, 43, 52–64.

- Condon, D. M. & Revelle, W. (2015a). Selected ICAR data from the SAPA-Project: Development and initial validation of a public-domain measure. *Harvard Dataverse*.
- Condon, D. M. & Revelle, W. (2015b). Selected personality data from the SAPA-Project: 08dec2013 to 26jul2014. *Harvard Dataverse*.
- Condon, D. M. & Revelle, W. (2015c). Selected personality data from the SAPA-Project: On the structure of phrased self-report items. *Journal of Open Psychology Data*, 3(1).
- Condon, D. M. & Revelle, W. (2016). Selected ICAR data from the SAPA-Project: Development and initial validation of a public-domain measure. *Journal of Open Psychology Data*.
- Costa, P. T. & McCrae, R. R. (1992). *NEO PI-R professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Costa, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the neo

- personality inventory. *Personality and Individual Differences*, 12(9), 887–898.
- Cureton, E. (1966). Corrected item-test correlations. *Psychometrika*, 31(1), 93–96.
- Deary, I. (2008). Why do intelligent people live longer? *Nature*, 456(7219), 175–176.
- DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the big five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, 33(4), 533–552.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the big five. *Journal of Personality and Social Psychology*, 93(5), 880–896.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417–440.
- Digman, J. M. (1997). Higher-order factors of the big five. *Journal of Personality and Social Psychology*, 73, 1246–1256.

- Dwyer, P. S. (1937). The determination of the factor loadings of a given test from the known factor loadings of other tests. *Psychometrika*, 2(3), 173–178.
- Eysenck, H. J. (1994). The big five or the giant three: Criteria for a paradigm. In C. F. Halverson, G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 37–51). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences*, 6(1), 21 – 29.
- French, J. A. & Condon, D. M. (2015). SAPA Tools: Tools to analyze the SAPA Project. R package version 0.1.
- Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on “estimating the reproducibility of psychological science”. *Science*, 351(6277), 1037.

- Goldberg, L. R. (1990). An alternative “description of personality”: The big-five factor structure. *Journal of Personality and Social Psychology*, 59(6), 1216–1229.
- Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, 4(1), 26–42.
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe*, volume 7 (pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.
- Goldberg, L. R. (2010). Personality, demographics and self reported acts: the development of avocational interest scales from estimates of the mount time spent in interest related activities. In C. Agnew, D. Carlston, W. Graziano, & J. Kelly (Eds.), *Then a miracle occurs: Focusing on the behavior in social psychological theory and research* (pp. 205–226). New York, NY: Oxford University Press.

- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79 – 132.
- Gough, H. G. & Bradley, P. (1996). Cpi manual . palo alto.
- Hayes, K. (1962). Genes, drives, and intellect. *Psychological Reports (Monograph Supplement 2)*, *10*, 299–342.
- Hendriks, A. A. J. (1997). *The construction of the five-factor personality inventory (FFPI)*. PhD thesis, Rijksuniversiteit Groningen, Groningen, The Netherlands.
- Hendriks, A. A. J., Hofstee, W. K., & De Raad, B. (1999). The five-factor personality inventory (FFPI). *Personality and Individual Differences*, *27*(2), 307 – 325.
- Hofstee, W. K., de Raad, B., & Goldberg, L. R. (1992). Integration of the big five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, *63*(1), 146–163.
- Hogan, R. & Hogan, J. (1995). *The Hogan personality inventory manual (2nd. ed.)*. Tulsa, OK: Hogan Assessment Systems.

- Horn, J. L. (1973). On extension analysis and its relation to correlations between variables and factor scores. *Multivariate Behavioral Research*, 8(4), 477 – 489.
- Jackson, D. N. (1983). *JPI-R: Jackson Personality Inventory, Revised*. Sigma Assessment Systems, Incorporated.
- Jackson, D. N., Paunonen, S. V., & Tremblay, P. F. (2000). *Six Factor Personality Questionnaire*. Port Huron, MI: Sigma Assessment Systems.
- Johnson, W. (2010). Extending and testing tom bouchard's experience producing drive theory. *Personality and Individual Differences*, 49(4), 296 – 301. Collected works from the Festschrift for Tom Bouchard, June 2009: A tribute to a vibrant scientific career.
- Kelly, E. L. & Fiske, D. W. (1950). The prediction of success in the VA training program in clinical psychology. *American Psychologist*, 5(8), 395 – 406.

- Kelly, E. L. & Fiske, D. W. (1951). *The prediction of performance in clinical psychology*. Ann Arbor, Michigan: University of Michigan Press.
- Kievit, R. A., Frankenhuys, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology*, 4(513), 1–14.
- Lee, K. & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39(2), 329–358.
- Liebert, M. (2006). A public-domain assessment of music preferences as a function of personality and general intelligence. Honors Thesis. Department of Psychology, Northwestern University.
- Lord, F. M. (1955). Estimating test reliability. *Educational and Psychological Measurement*, 15, 325–336.
- Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The electronically activated recorder (ear): A

device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers*, 33(4), 517–523.

- Mehl, M. R. & Robbins, M. L. (2012). Naturalistic observation sampling: The electronically activated recorder (ear). In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life*. New York, NY: Guilford Press.
- Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, 317(5834), 82.
- Mottus, R. (2016). Towards more rigorous personality trait-outcome research. *European Journal of Personality*.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factors structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 66(6), 574–583.

- ottus, R. M., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2016). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, (in press).
- Pozzebon, J. A., Visser, B. A., Ashton, M. C., Lee, K., & Goldberg, L. R. (2010). Psychometric characteristics of a public-domain self-report measure of vocational interests: The oregon vocational interest scales. *Journal of Personality Assessment*, 92(2), 168–174.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rafaeli, E., Rogers, G. M., & Revelle, W. (2007). Affective synchrony: Individual differences in mixed emotions. *Personality and Social Psychology Bulletin*, 33(7), 915–932.
- Revelle, W. (2015). *psych: Procedures for Personality and*

Psychological Research.

<http://cran.r-project.org/web/packages/psych/>: Northwestern University, Evanston. R package version 1.5.8.

- Revelle, W. & Condon, D. M. (2012). Multilevel analysis of personality: Personality of college majors. Presented at the annual meeting of the Society of Multivariate Experimental Psychology.
- Revelle, W. & Condon, D. M. (2015a). Ability, temperament, and interests: their joint predictive power for job choice. International Society for the Study of Intelligence. Albuquerque, New Mexico.
- Revelle, W. & Condon, D. M. (2015b). A model for personality at three levels. *Journal of Research in Personality*, 56, 70–81.
- Revelle, W., Condon, D. M., Wilt, J., French, J. A., Brown, A., & Elleman, L. G. (2016). Web and phone based data collection using planned missing designs. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *SAGE Handbook of Online Research Methods* chapter 37: Mobile Methods. Sage Publications, Inc. (in press).

- Revelle, W. & Wilt, J. (2015). The data box and within subject analyses: A comment on Nesselroade and Molenaar. *Multivariate Behavioral Research (in press)*.
- Revelle, W., Wilt, J., & Condon, D. (2011). Individual differences and differential psychology: A brief history and prospect. In T. Chamorro-Premuzic, A. Furnham, & S. von Stumm (Eds.), *Handbook of Individual Differences* chapter 1, (pp. 3–38). Oxford: Wiley-Blackwell.
- Revelle, W., Wilt, J., & Rosenthal, A. (2010). Individual differences in cognition: New methods for examining the personality-cognition link. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of Individual Differences in Cognition: Attention, Memory and Executive Control* chapter 2, (pp. 27–49). New York, N.Y.: Springer.
- Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability

for predicting important life outcomes. *Perspectives on Psychological Science*, 2(4), 313–345.

Rounds, J., Su, R., Lewis, P., & Rivkin, D. (2010). O* NET® interest profiler short form psychometric characteristics: Summary.

Shibley, W. C. (2009). *Shibley-2: manual*. Western Psychological Services.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 13(2), 238–241.

Tellegen, A. & Waller, N. G. (2008). Exploring personality through test construction: Development of the multidimensional personality questionnaire. *The Sage handbook of personality theory and assessment*, 2, 261–292.

Tupes, E. C. & Christal, R. E. (1961). Recurrent personality factors based on trait ratings. Technical Report 61-97, USAF ASD

Technical Report, Lackland Air Force Base (reprinted in *Journal of Personality* (1992) 60: 225–251).

Wilt, J. (2014). *A new form and function for personality*. PhD thesis, Northwestern University.

Wilt, J., Bleidorn, W., & Revelle, W. (2016). Finding a life worth living: Meaning in life and graduation from college. *European Journal of Personality*, (in press).

Wilt, J., Funkhouser, K., & Revelle, W. (2011). The dynamic relationships of affective synchrony to perceptions of situations. *Journal of Research in Personality*, 45, 309–321.

Wilt, J. & Revelle, W. (2015). Affect, behaviour, cognition and desire in the big five: An analysis of item content and structure. *European Journal of Personality*, 29(4), 478–497.

Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika*, 2(2), 121–134.