Science by analogy: PWAS or Persome Wide Association Studies

Presented to the Kellogg Behavioral Brown Bag

William Revelle

Northwestern University
Evanston, Illinois USA

Slides available at personality-project.org/sapa
R code included as an appendix

October 16, 2020

Outline

Overview

Open Science

- Astronomy

Radio Astronomy:Synthetic Aperture Telescopes::Synthetic Aperture Personality Assessment (SAPA):Personality

- Measuring individual differences: the tradeoff between breadth versus depth

Items, not latent traits: The utility of using lots of items

Genome Wide Association Studies:GWAS::Persome Wide Association Studies: PWAS

- Profiles

Big Data

Summary

R code for analyses

- Replicate on a much larger data set.

Open Science: A new idea or a long term tradition?

1. Science is a process for asking questions that have answers
    - Our questions and our answers need to be open and shared.
    - Our way of addressing these questions should be open to others.
    - Our results are for everyone, not just those who can afford to pay for journals.
    - Our results need to trusted and trustworthy.
2. This is not a new idea, sharing ideas, methods and results is as old as the Royal Society from 1660.
    - It was an "invisible college" of natural philosophers and physicians.
    - Royal Society's motto 'Nullius in verba' is taken to mean 'take nobody's word for it'. (We might now say, does it replicate?)
3. Personality research is an example of open science.
    - Tends to be well powered and replicable.
    - Tends to involve multiple studies over multiple years.
    - Growing tendency to use open and shared materials.

Questions we ask in personality

1. Kluckholm and Murray's (Kluckhohn & Murray, 1948) basic trichotomy remains active today (Revelle, 1995).
   - All people are the same (human nature)
   - Some people are the same (individual differences)
   - No person is the same (unique life stories of the individual)
2. Much of personality research is at this middle level of how some people are the same and differ from other people.
   - Description of individual differences
     - Dimensional models include Block's 2 (Block, 1971, 2002), Eysenck's Giant 3 (Eysenck, 1994), Big 5 (Digman & Takemoto-Chock, 1981; Digman, 1990; Goldberg, 1990), 8-9 (Comrey, 1995), Cattell's 16 (Cattell & Stice, 1957), and even Condon's "little 27" (Condon, 2018)
   - Different theoretical explanations of individual differences
     - SocioAnalytic (Hogan, 1982; Hogan & Blickle, 2018)
     - Biological (Eysenck, 1967; Gray, 1991; Corr, 2002; DeYoung, 2010, 2015)
   - Practical use of individual differences
     - Prediction of leadership effectiveness (Hogan, 2007), academic performance (Sackett & Kuncel, 2018) mortality, marital status, occupational choice, and mental health (Ozer & Benet-Martinez, 2006).

Traditional latent trait approach to measurement of personality

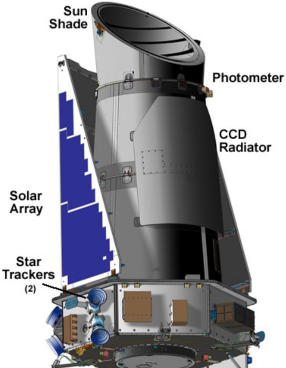
1. Known since Spearman (1904) that measures are befuddled with error.
2. Can reduce befuddlement (increase reliability) by aggregating items (Brown, 1910; Spearman, 1910).
3. Structure of scales can be analyzed by latent trait (factor analytic) or components (not latent trait models, but frequently confused with them).
4. Factor analytic approaches led to convergence on a "consensual structure" of 5 factors (Digman, 1990; Goldberg, 1990),
5. Then, there became a race to bottom in developing shorter and shorter measures of the Big 5.
    - Costa and McCrae's 300 items of the NEO-PI (Costa & McCrae, 1992)
    - Goldberg's original set of 100 adjectives (Goldberg, 1992)
    - Saucier and the 40 mini markers Saucier (1994) and Oliver John et al (John, Donahue & Kentle, 1991) 44 phrased items.
    - Rammstedt and John's 10 items (Rammstedt & John, 2007) and the Gosling et al TIPI (Gosling, Rentfrow & Swann, 2003).
    - The lower bound: the 5 items of Konstabel (Konstabel, Lönnqvist, Leikas, Velàzquez, H, Verkasalo, & et al., 2017).

A different approach: the power of the item

1. But personality $\neq$ Big 5.
2. An alternative approach to giving fewer and fewer items to measure just the Big 5 is to give more and more items to measure as much of the personality domain as possible.
3. My colleagues and I are now examining the structure of more than 6,000 items and are on the way to 10,000 (Condon, 2018; Revelle, Wilt & Rosenthal, 2010; Revelle, Condon, Wilt, French, Brown & Elleman, 2016).
4. We do this because we think that although only about 20% of any item measures a single higher order trait, at least 80-90% of an item is reliable variance.
5. We need ways to give more items and to examine the total reliable variance of the item.
6. But how to do this?
7. By apply techniques analogous to those of radio astronomy but already known to psychologists (Lord, 1955b).

A now for something completely different: astronomy

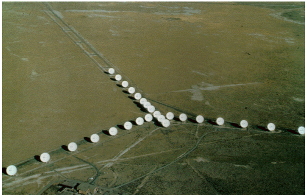
Resolution varies by aperture diameter (bigger is better)

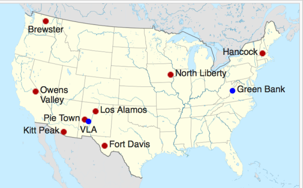


A short diversion: history of radio telescopes

Just as with optical 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 (subjects) and n (items) at the same time?

1. Frederic Lord (1955a) 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.
8. My colleagues and I have discussed this technique for several years as a way of embracing your missingness (Revelle et al., 2010, 2016).

The basic problem: Fidelity versus bandwidth

1. Many personality traits, interests and cognitive abilities are multidimensional and have complex structure.
    - To measure these, we need to have the precision that comes with many participants.
    - But we also need the bandwidth that comes with many items.
    - But participants are reluctant to answer very many items.
2. This has led to the quandary of should you give many people a few items or a few people, many items?
3. Our answer is to do both, but with a *Massively Missing Completely At Random* (MMCAR) data structure.
4. We refer to this technique as *Synthetic Aperture Personality Assessment* (SAPA) to recognize the analogy to synthetic aperture radio astronomy (Revelle et al., 2010, 2016).
5. This is functionally what Frederic Lord (1955a, 1977) suggested 65 years ago. It is time to take him seriously.

SAPA overview

1. At the sapa-project.org we use Synthetic Aperture Personality Assessment (SAPA) methods to assess $\approx 30K$ participants per month. This is just a technique of Massively Missing Completely at Random (MMCAR) data presentation. Each participant is given a random subset of items chosen from an item pool of more than 6600 items. These items, extended from the International Personality Item Pool (Goldberg, 1999) and the International Cognitive Ability Resource (Condon & Revelle, 2014; Revelle, Dworak & Condon, 2020a), assess temperament, cognitive ability, interests and attitudes as well as self reported behaviors and demographic information.
2. Conventional psychometric techniques (both classical and IRT) are used to identify homogeneous scales; empirical item selection procedures are use to develop optimal item composites to predict a wide range of criteria. Data analysis code is done using the *psych* package (Revelle, 2020) in R (R Core Team,

Lord (1955a) and matrix sampling

1. Given an N (subjects) by n (item) matrix, we can sample:
2. Type 1: Subjects – basic statistical theory
   - $\bar{x}$ and its standard error $\sqrt{\frac{\sigma^2}{N-1}}$
   - r_{xy} and its standard error $\sqrt{\frac{1-r^2}{N-2}}$
3. Type 2: Items – this is the basis of classical reliability theory especially domain sampling (Tryon, 1957, 1959):
   - $KR_{20} = \alpha = \lambda_3$ represent the correlation of a test with a test just like it sampled from a larger population of items.
   - ω_h and ω_t similarly are estimates of what the general factor, ω_h, or total, ω_t, correlation would be with another representation in the domain. (See Revelle & Condon, 2019, for everything you want to know about reliability but were afraid to ask).
4. Type 12: Matrix sampling of subjects and items
   - Special case is balanced incomplete blocks (BIB).
   - General case is Missing Completely at Random (MCAR).

3 Methods of collecting 256 subject * items data

1) 8 x 32 complete
```
46213634521143453443645331212414
21243623166421516154432261516513
51661351155165463622224435623344
11141343362332215612152135614522
25353121264561433433232246526411
61335154566424114612641225353516
24634342151536242425413513435116
11554654453123111162423325516334

Type 1 = sample subjects

Type 2 = sample items

Type 12 sample items and subjects
```

2) 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
```

12) 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...
```

3 Methods of collecting 256 subject * items data

```
1) complete (Ideal)                   2) Sample people                 3) Items
22552141414336514122645166143244      22552141414336514122645166143244 22552141
32144265454235634562343524256611      ................................ 32144265
43553143152141541641526114551151      ................................ 43553143
52654223445614444431162645313124      ................................ 52654223
62222255242315442652355414213325      ................................ 62222255
22125412454242154221456444214564      ................................ 22125412
65113311244511226522615346451412      ................................ 65113311
54436452425245244554632246526466      ................................ 54436452
55223643555215245514633426121226      55223643555215245514633426121226 55223643
35522554332664265346655451531612      ................................ 35522554
63261241341466311243222233323541      63261241341466311243222233323541 63261241
32224431433144451645255464435552      ................................ 32224431
11564655513111334341463561655541      11564655513111334341463561655541 11564655
24532624664444656366642463322555      ................................ 24532624
25516362264523255665245644125611      ................................ 25516362
32255635422342631523143414221354      ................................ 32255635
23244456631411361161615126144214      ................................ 23244456
34526633236542563633625123624421      ................................ 34526633
13451522616451531355135621451536      13451522616451531355135621451536 13451522
31625444241623135123121345134162      ................................ 31625444
44252526365556663522524162313453      ................................ 44252526
54361436651313615433261662235132      ................................ 54361436
46635454552135645224352362433436      46635454552135645224352362433436 46635454
26511624245416441145655363265265      ................................ 26511624
63512331235542645524352562623235      ................................ 63512331
11523665433656446452523322216333      11523665433656446452523322216333 11523665
56436532623253433145633663651242      ................................ 56436532
15136366233651513351113353151452      ................................ 15136366
46321152211446344326554442255226      ................................ 46321152
62156523111352364233551656146433      ................................ 62156523
```

NORTHWESTERN UNIVERSITY

12 (Matrix) Sampling Methods of collecting 256 subject * items data

a) 32 x 16 balanced incomplete

```
................4122645166143244
................4562343524256611
................1641526114551151
................4431162645313124
................2652355414213325
........45424215........44214564
........24451122........46451412
........42524524........46526466
........55521524........26121226
........33266426........51531612
........3414663112432222........
........4331444516452554........
........5131113343414635........
........6644446563666424........
........2645232556652456........
32255635................14221354
23244456................26144214
34526633................23624421
13451522................21451536
31625444................45134162
44252526........35225241........
54361436........54332616........
46635454........52243523........
26511624........11456553........
63512331........55243525........
1152366543365644................
5643653262325343................
1513636623365151................
4632115221144634................
6215652311135236................
```

b) 32 x 8 SAPA p =.25

```
..55.....1......4...6..16.....4.
.2..4...45.....3........2.2....1
..5.....1......4...1..6...551...
..6..2.........4.......64...3.24
..22..5.......4.2..2...4..2.....
.2.2..1.4.....1..2......4.....6.
......1124......65...1...6......
.44..4.2......2.......2...52....
.5..3...5......4..1.6.......1..6
....25....2....6....65.......61.
....1...3.....311.4..22.........
...244......44....45...........2
.1.......1.11...4.4....5......4.
.....6.4.........3.66.2....2..5.
25...3.2.........6......4.12....
......3..22....3....1....42..3..
..2...5..31..1..........2....21.
...2.6...3...2.6.......12..2....
.....5...1...1.3135.1...........
...2..4........35.........13.16.
..2....6..5.......2....1....3.53
543.............5...2.....2.51..
4...54....2...6............3..36
....1..424....4.....6.......5.6.
........235...6..........262...5
.15....5..3....4.4........2..3..
........6.3......1.5.3..63...2..
..1...66.........35..1.35.......
......5221......4.......42..5...
.21........3..........1.5.1.6..3
```

NORTHWESTERN UNIVERSITY

Type 12 sampling (matrix sampling)

1. Balanced incomplete blocks works but is hard if giving less than 50% coverage
   - 50% requires 6 blocks to be fully balanced (divide into 4ths and then present all pairs of the fourths):
     - AB, AC, AD, BC, BD, CD where A, B, C, and D are 1/4 of the total.
   - Even then, items within blocks co-occur three times as much as items between blocks.
   - 33% samples require 15 blocks, 25% 28 blocks.
2. SAPA sampling (Massively Missing Completely at Random) allows any sampling rate.
3. BIB can be done with printed forms, MMCAR requires computer administration.
4. Possible to do FIML with BIB design, need to do pairwise complete for SAPA. But, because it is MMCAR, it is unbiased.

Why we care: Breadth vs. depth of measurement

1. Factor structure of domains needs multiple constructs to define structure.
2. Each construct needs multiple items to be measured reliably.
3. This leads to an explosion of potential items.
4. But, people are willing to answer only a limited number of items.
5. This leads to the use of short and shorter forms (the NEO-PI-R (Costa & McCrae, 1992) with 300, the IPIP (Goldberg, 1999) Big 5 with 100, the BFI (John et al., 1991) with 44 items, the BFI2 (Soto & John, 2017) with 60, the 30 item 'Short Five' (Konstabel et al., 2017), the TIPI (Gosling et al., 2003) with 10 and the 10 item BFI (Rammstedt & John, 2007)) to include as part of other surveys.
6. Unfortunately, with this reduction of items, breadth of substantive content is lost. We offer an alternative procedure.

Subjects are expensive, so are items

1. In a survey such as Amazon's Mechanical Turk (MTurk), we would need to pay by the person and by the item.
2. Volunteer subjects are not very willing to answer many items.
3. Why give each person the same items? Sample items, as we sample people (Lord, 1955b, 1977).
4. 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) (we sometimes have 99% missing data although our median is only 93% missing!)
5. This is the essence of Synthetic Aperture Personality Assessment (SAPA) (Condon & Revelle, 2014; Condon, 2014; Revelle et al., 2016, 2010).
6. This is a much higher rate of missingness than discussed in the balanced incomplete block design of NAEPS or PISA.

3 Methods of collecting 256 subject * items data

a) 8 x 32 complete

```
46213634521143453443645331212414
21243623166421516154432261516513
51661351155165463622224435623344
11141343362332215612152135614522
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
```

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..
```

Synthetic Aperture Personality Assessment

1. Give each participant a random sample of *pn* items taken from a larger pool of n items. p_i might be anywhere from .01 to 1.
2. Find covariances based upon "pairwise complete data". Each pair appears with probability $p_i p_j$ with a median of .01.
3. Find scales based upon basic covariance algebra.
   - Let the raw data be the matrix ${}_N\boldsymbol{X}_n$ with N observations converted to deviation scores for n items.
   - Then the item variance covariance matrix is ${}_n\boldsymbol{C}_n = \boldsymbol{X}'\boldsymbol{X}N^{-1}$
   - and scale scores, ${}_N\boldsymbol{S}_s$ are found by $\boldsymbol{S} = {}_N\boldsymbol{X}_{nn}\boldsymbol{K}_s$.
   - ${}_n\boldsymbol{K}_s$ is a keying matrix, with $k_{ij} = 1$ 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,
     ${}_s\boldsymbol{C}_s = {}_s\boldsymbol{S}'_{NN}\boldsymbol{S}_s N^{-1} =$
     $$ {}_s\boldsymbol{C}_s = (\boldsymbol{X}\boldsymbol{K})'(\boldsymbol{X}\boldsymbol{K})N^{-1} = \boldsymbol{K}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{K}N^{-1} = \boldsymbol{K}'{}_n\boldsymbol{C}_n\boldsymbol{K}. \quad (1)$$
4. That is, we can find the correlations/covariances between scales from the item covariances, not the raw items.

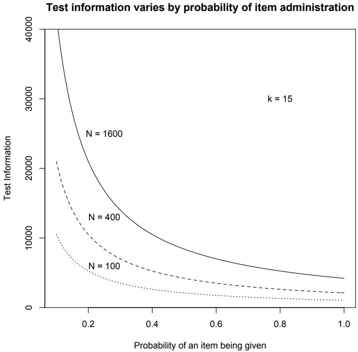
Total information

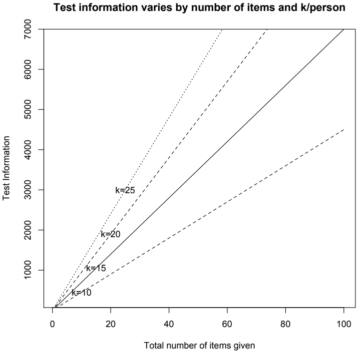
1. The information in a single correlation varies by the reciprocal of its standard error $\sigma_r = \sqrt{\frac{1-r^2}{N-2}}$ or $I = \sqrt{\frac{N-2}{1-r^2}}$
2. In SAPA, k items/person are randomly selected with probability p from a larger number, n of items (k = pn).
3. Thus,the number of subjects per item is pN.
4. The total number of correlations is just $\frac{n*(n-1)}{2}$ and the number of subjects per correlation is $p^2 N$.
5. Total information is number of correlations * $\sqrt{p^2 N} = \frac{n*(n-1)}{2}\sqrt{p^2 N} = \frac{(k/p)((k/p)-1)}{2} * \sqrt{p^2 N} = \frac{k*(k-1)\sqrt{N}}{2*p}$.
6. For the "normal case" where $p = 1$, the information is just what we expect–a quadratic function of k: $I_{kN} = \frac{k*(k-1)\sqrt{N}}{2}$.
7. But the more interesting case (the SAPA case) is for $p < 1$ the information is a hyperbolic function of p: $I_{pkN} = \frac{k*(k-1)\sqrt{N}}{2*p}$ but a linear function of the total number of items given (n= k/p) $I_{pkN} = \frac{n*(k-1)}{2} * \sqrt{N}$

Total information varies by the number of items (n) and the probability of sampling (p) and total sample size (N)

For k items/subject and N subjects, if every item is given with probability p, the information in the test is

$$I_{pkN} = \frac{k*(k-1)\sqrt{N}}{2*p} = \frac{n*(k-1)}{2} * \sqrt{N}$$





Theoretical demonstrations show this technique works with as few as 200 subjects

1. We have shown demonstrations of this technique for sampling from 10,000s of subjects (Revelle et al., 2010, 2016) with real data.
2. David Condon and I have reported on simulations of factor recovery with 1,000s of subjects (Revelle & Condon, 2017; Revelle, 2019).
3. Sonja Heintz at the University of Geneva, Elizabeth Dworak at NU, David Condon (University of Oregon) and I have shown this technique works for as few as 200 subjects and can be applied to ESM data (Revelle, Condon & Heintz, 2018).
4. Our empirical investigations was originally based upon the open source International Personality Item Pool.

How do we get subjects?

1. Use the web and give feedback.
2. People like to be told about themselves.
3. The `outofservice.com` web site used by Sam Gosling, the Facebook site used by Kozinski and Stillwell, the site used by Soto, all of these work.
4. We have our own site where we emphasize sampling of items (sapa-project.org).
    - Originally hosted on a Mac in my lab
    - Moved to NUIT for better service after PMC lab serve was hacked.
    - Moved to Google Cloud for even better service when demand peaked at 50-60K /day one day.
5. The feedback is based upon a 5 and 27 factor model.

IPIP and the personality assessment

1. Lew Goldbergs's International Personality Item Pool (IPIP) was very controversial when first released (Goldberg, 1999) but has helped establish the common measurement of personality by creating and administering short item stems that capture the essence of most published personality inventories.
2. Goldberg and his colleagues at the University of Oregon developed the Eugene-Springfield sample (Goldberg & Saucier, 2016) which has given several thousand items to $\approx 1,000$ predominantly white middle class participants over 10 years. This sample has been the basis of the development and validation of the International Personality Item Pool (see ipip.ori.org).
3. In fact, many of the subsequent attempts at personality scale development have used the Eugene-Springfield sample, e.g., the BFI-2 (Soto & John, 2017), and the Big Five Aspect Scales (BFAS) of DeYoung, Quilty & Peterson (2007).

The Eugene Springfield sample and the IPIP are WEIRD

1. Unfortunately, many of the items that have come out of the E-S sample were prematurely selected to represent the Big 5. That is, even though meant to capture the many dimensions of the lexicon, the adjectival descriptors used had been trimmed to those matching the 5 factors that have been known since the 1950's (Kelly & Fiske, 1950, 1951; Tupes & Christal, 1961; Norman, 1963).
2. Because of the ease of use and the openness of the IPIP, most of the short forms followed the Big Five structure that came out of the E-S sample.
3. SAPA subjects are less WEIRD, but still not typical.

Characteristics of SAPA reported here

1. Total number in shared data sets discussed today 126,884. Something great than 1,000,000 total participants have been collected since 2010.
2. Age 14-90 (mean = 26, median= 22)
3. Gender 63% Female (have switched to non-binary scale for more recent participants)
4. Education 15% less than 12 years, 9% HS grad, 41% in college, 6% some college 15% BA, 5% in grad school, 10% Grad or prof degree
5. 68% US, 4.3% Can, 3.7% UK, 2.1% AUS, ...

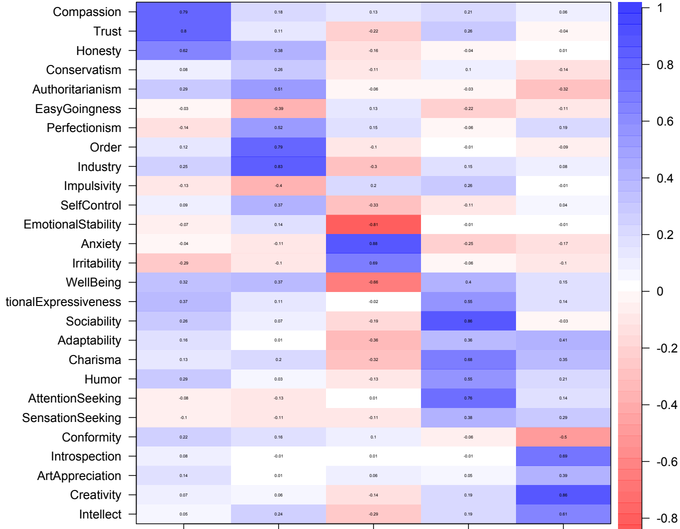
More items, alternative structures

1. Of about 2,084 item in the IPIP, representing 200 different scales, David Condon found that 696 items were actually unique and had no dominant factor structure (Condon, 2018) . However, he found that 135 of the items could be well organized in terms of 5 broad factors (the Big 5) and 27 narrower factors (the little 27).
2. Scores for 4,000 visitors to the SAPA-project for these 135 items and 10 criteria are included in the *psychTools* package which accompanies the *psych* package (Revelle, 2020) for R (R Core Team, 2020).
3. I am going to use this example set for a series of demonstrations. To encourage you to do these analyses yourself, I include the R code as an appendix.
4. I will also discuss another public data set for 126,884 participants with scores on the 696 items and 22 distinct criteria (Condon & Revelle, 2015; Condon, Roney & Revelle, 2017b,a).

More items leads to improved measures at multiple levels

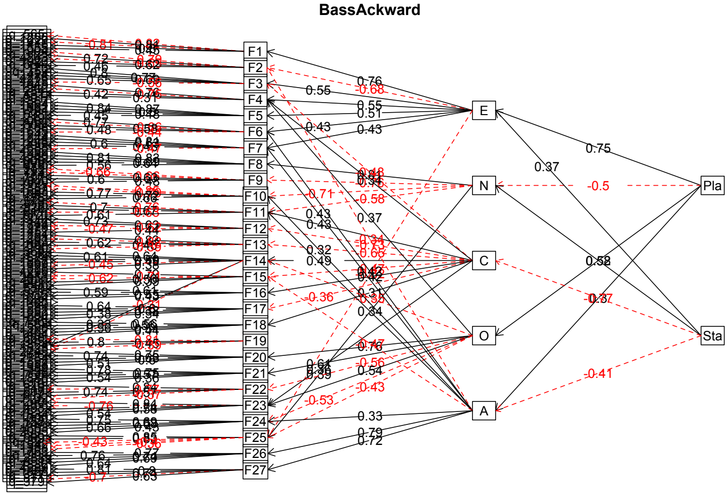
1. Better reliability of high level traits (e.g., Big 5)
    - The Big 5 scales from the `spi` are 14 item scales with an average α of .87 with a mean ω_h of .67 and ω_t of .91.
    - The little 27 are five items scales with mean α of .82 (ω_h is not really interpretable for item scales).
2. The little 27 are not meant to be facets of the big 5 but are rather narrower constructs.
3. This is best shown graphically as a `corPlot` and a `bassAckward` plot.

The structure of the spi is both 5 and 27 factors



NORTHWESTERN UNIVERSITY

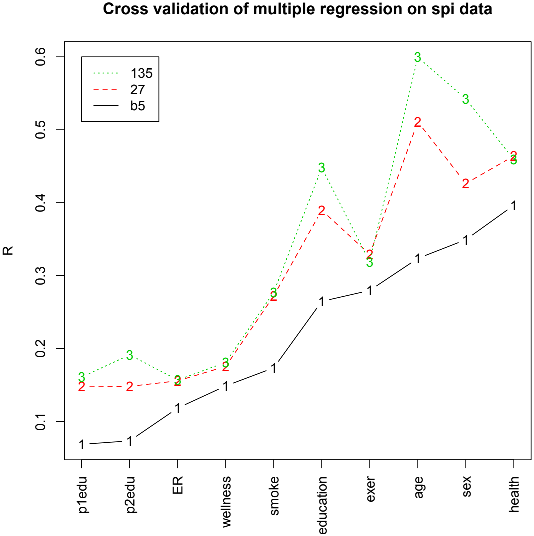
bassAckward of the 135 spi items with 2 - 5 and 27 factor solutions



What about prediction?

1. We have examined structure, but how useful are these various levels of analysis?
2. Multiple regression of 10 criteria from the Big 5, the little 27, and the items.
3. One would expect to overfit the data if we use so many predictors, thus, we need to apply cross validation.
4. For some analyses, (e.g. `bestScales` we use "bagging" (boot strap aggregation) or "kfolds". Here we just do normal cross validation.
5. Derive model on half the sample, cross validate on the other half.
6. Plot the results.

Cross validation for 5, 27 and 135 predictors for the `spi`



1. Criteria differ in predictability
2. 135 items better than 27 factors
3. 27 better than 5

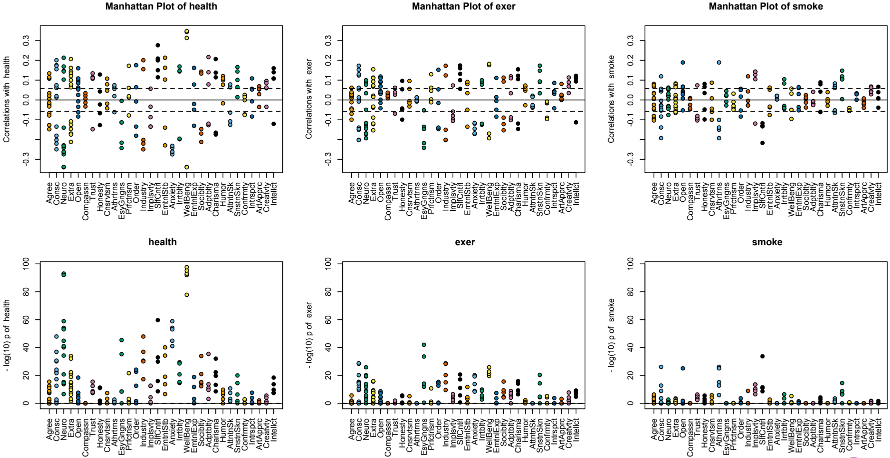
Yet another analogy: genetics

1. Most target gene studies have been dreadfully underpowered and produce too many type I errors.
2. With the exception of a few genes (color blindness, PKU), most genetic effects are very small.
3. Each Single Nucleotide Polymorphism (SNP) accounts for very little variance.
4. But with the ability to do Genome Wide studies aggregated across 100,000s to 1,000,000s of people, it is now possible to reliably identify SNPS associated with phenotypic traits.
5. It is also possible to find genetic propensity scores (basically just linear sums) of 1,000s SNPs at a time.
6. GWAS also introduces the concept of a genetic correlation, which is the correlation across the genome of effect sizes.
7. These genetic correlation assess the amount that the genetic variance in any two phenotypes is similar.

Analogous to GWAS is Persome Wide Association Studies (PWAS)

1. "Manhattan" plots are just ways of displaying GWAS or PWAS correlations.
2. In GWAS the plots are SNPS by chromosome.
3. in PWAS we organize the items by the scale they are associated with.
4. We do this for the `spi` data on three criteria: health, exercise and smoking.
5. In the following figure I show the correlation of 135 items with each of the three criteria. I group the items by their scales.
6. I also show - the log of the probability of the correlation.

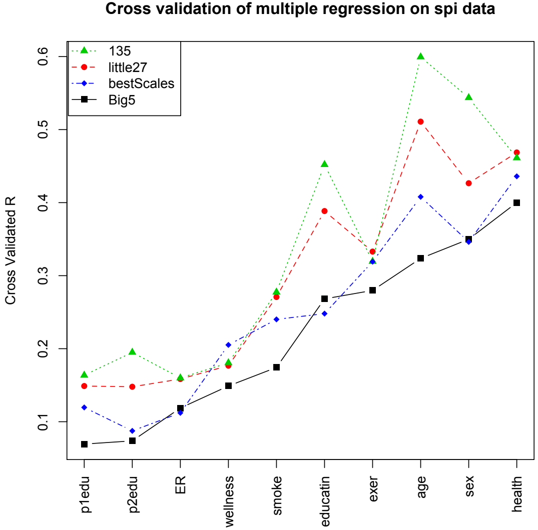
Manhattan plots can show the raw correlations or -log p values



An alternative to regression: `bestScales`

1. An alternative to multiple regression is to choose the best unit weighted items. (see the Manhattan plots)
2. We describe a new algorithm (BISCUIT aka `bestScales`) based upon very old ideas (Elleman, McDougald, Revelle & Condon, 2020).
3. Choose items most correlated with a criterion. Cross validate these multiple times (using kfolds or bagging) and then form the unit weighted composites.
4. Based upon the "Robust beauty of improper linear models" (Dawes, 1979) and the idea that regression weights are fungible (Waller, 2008).
5. Generally pretty good, if not optimal, and much more understandable in that we can examine what the best items are.
6. We do this for the `spi` data set and compare the cross validated correlations with those of the Big5, little 27 and 135 item multiple Rs.

Cross validation for 5, 27, 135 and bestScalesfor the `spi`



1. Best scales (made up of the top 20 items are not as good as
2. linear regression from all 135 items
3. linear regression from 27 factors (using 135 items)
4. but are better than big 5 (using 70 items)

What are the best items predicting these criteria?

Table: Smoking

A table from the psych package in R

Variable	Freq	men.r	sd.r	item
q_1461	10	-0.24	0.01	Never spend more than I can afford.
q_1867	10	-0.20	0.01	Try to follow the rules.
q_1609	10	0.19	0.01	Rebel against authority.
q_1173	10	0.17	0.01	Jump into things without thinking.
q_1624	10	-0.17	0.01	Respect authority.
q_369	10	-0.16	0.01	Believe that laws should be strictly enforced.
q_56	10	-0.16	0.01	Am able to control my cravings.
q_35	10	0.16	0.01	Act without thinking.
q_1462	10	-0.15	0.01	Never splurge.
q_1424	10	0.15	0.01	Make rash decisions.
q_736	10	-0.15	0.01	Easily resist temptations.
q_598	10	0.14	0.01	Do crazy things.
q_1590	10	-0.13	0.01	Rarely overindulge.
q_1452	9	0.13	0.01	Neglect my duties.
q_4276	9	0.12	0.01	Often make decisions on the spur of the moment.

Best items predicting rated health

Table: health

A table from the psych package in R

Variable	Freq	men.r	sd.r	item
q_820	10	0.36	0.01	Feel comfortable with myself.
q_811	10	-0.35	0.01	Feel a sense of worthlessness /hopelessness.
q_2765	10	0.35	0.00	Am happy with my life.
q_578	10	-0.34	0.01	Dislike myself.
q_1371	10	0.31	0.01	Love life.
q_56	10	0.28	0.01	Am able to control my cravings.
q_1505	10	-0.28	0.01	Panic easily.
q_808	10	-0.27	0.01	Fear for the worst.
q_4249	10	-0.27	0.01	Would call myself a nervous person.
q_1452	10	-0.24	0.01	Neglect my duties.
q_979	10	-0.24	0.01	Get overwhelmed by emotions.
q_39	10	0.24	0.01	Adjust easily.
q_4252	10	-0.24	0.01	Am a worrier.
q_1444	10	-0.23	0.01	Need a push to get started.
q_1024	10	-0.23	0.01	Hang around doing nothing.
q_1840	10	0.23	0.01	my moods don't change more than most peoples.

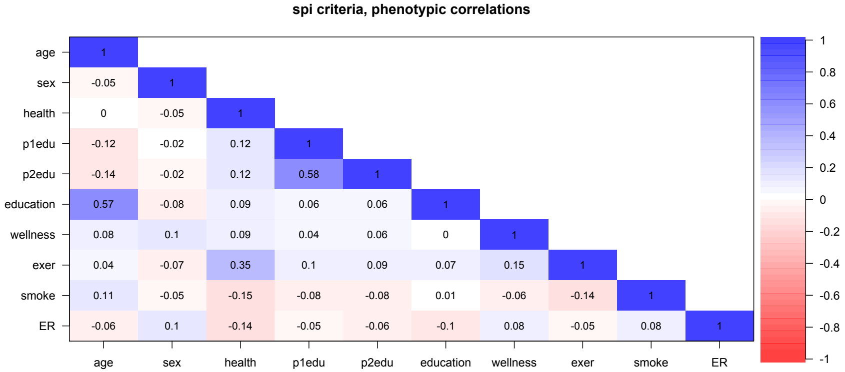
bestScales are analogous to GPS

1. in genetics, based upon the GWAS correlations, one forms *Genetic Propensity Scores* (GPS).
2. GPS are derived from large samples but can then be applied in smaller samples.
3. Similar to GPS from GWAS we have Personality Propensity Scores (PPS)
4. We can derived PPS from large samples (to increase the power and minimize the effect of chance) and then apply these PPS to smaller samples.

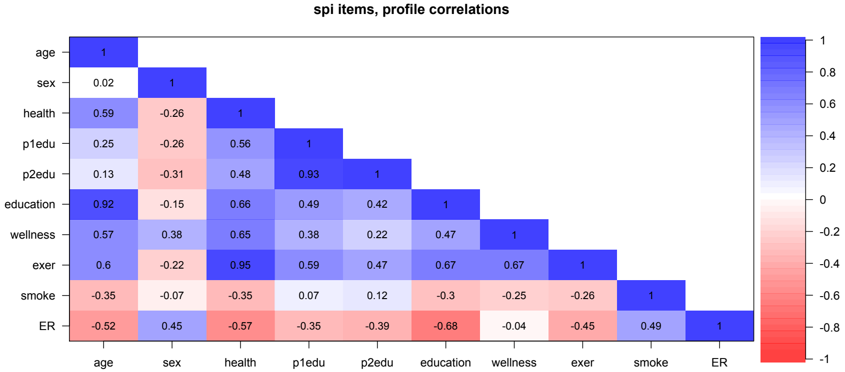
PWAS correlations

1. Genetic correlations are correlations taken across the genome and reflect the amount of shared genetic variance in two pheontypes.
2. So, we can find the profile correlation across the persome to examine shared predictable variance of phenotypes
3. I show three different correlation plots
   - Phenotypic correlations of our 10 `spi` crtieria
   - Profile correlations of these same 10 criteria where the profile is essentially the Manhattan plot
   - To compare these two, I combine them into one plot

Phenotypic correlations of the spi criteria

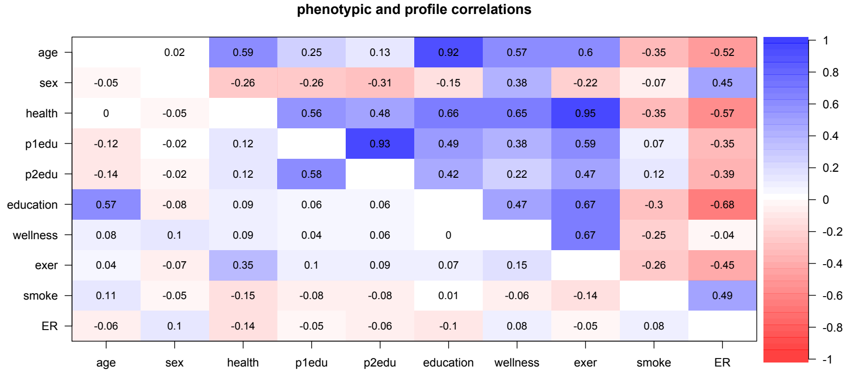


Profile correlations of the spi criteria



Show both the phenotypic and profile correlations

Compare the magnitude of the effects



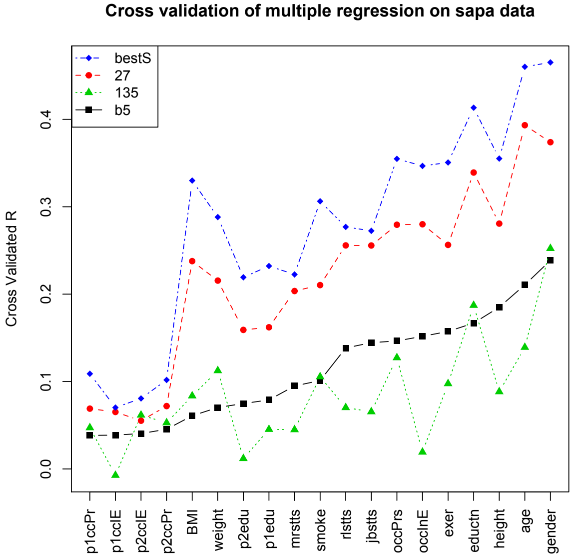
Profile correlations reflect shared predictable variance

1. Phenotypic correlations reflect all of the variance of the criteria.
2. Profile correlations reflect shared *predictable* variance.
3. Do we achieve a better understanding of the phenomena by examining what they have in common?
4. Consider the correlation between exercise and health (.35 verus .95), Emegency Room visits and smoking (.08 versus .49)
5. Is this an alternative way to adjust correlations for reliability?
6. NO.

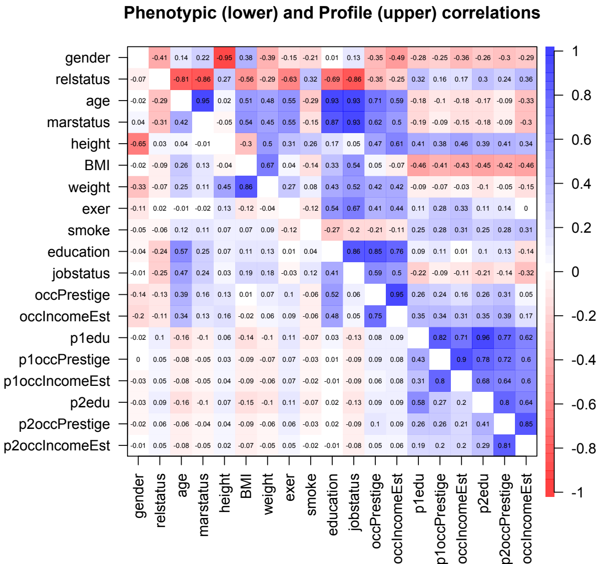
We can replicate this with 126,884 cases

1. The data are taken from DataVerse Condon & Revelle (2015); Condon et al. (2017a,b)
2. These data are discussed in two recent articles: Revelle, Dworak & Condon (2020b) and Dworak & Revelle (2020)
3. I show just a few analyses from those articles.
4. First the cross validated prediction
5. Then the profile results.

Comparing Big 5, little 27, 135 item regressions with best of 696



19 criteria phenotypic versus profile correlations



Profiles have more uses than shown

1. Profile weights can be derived for one criteria but can predict many more.
2. We have previously shown that the profile technique can be used to cluster the similarities of countries based upon the personality profiles that best predict dummy code country
3. We are doing this for college major and for occupations. By definition, majors are distinct and the phenotypic correlations will be slightly negative, but the profiles show how the natural sciences differ from the social sciences.
4. Even if you have just the 44 items of the BFI or the 60 of the BFI-2, these profile techniques can be applied to your data as well.

Conclusion and an invitation

1. Other sciences have developed techniques that we can share (at least by analogy).
2. Combining techniques similar to those from Radio Astronomy and from genetics allows us to ask different questions than we have been asking.
3. Items have much more information that we think (although the developers of empirical methods such as Gough (1957) or Hathaway & McKinley (1943) knew this years ago).
4. It is time to rethink our reliance on latent variable models. Perhaps we should focus on observables that we care about.
5. This is a direct challenge to those of us who like to think in casual models and the biological basis of personality.
6. Am I advocating personality engineering or personality theory? I am not sure.
7. However, I am sure that it might be time for us to rethink our reliance on latent trait models.

Need for open science

1. These techniques rely on shared materials, shared methods, and open science.
2. Can we use SAPA like techniques to refocus on the power of the item and move beyond the Big 5?
3. We have used a similar approach in the measurement of ability in the International Cognitive Ability Resource (ICAR) (Condon & Revelle, 2014; Condon, Doebler, Holling, Gühne, Rust, Stillwell, Sun, Chan, Loe & Revelle, 2014; Dworak, Revelle, Doebler & Condon, 2020; Revelle et al., 2020a).
4. By combining traditional temperament measures (e.g. the `spi` items or the magic 696 with measures of interests and ability, we can go even further.
5. Join us.

Slides, data and code are available for all to use

1. This work reflects contributions from David Condon, Liz Dworak, Lorien Elleman and members of the Personality, Motivation and Cognition Laboratory (aka the Telemetrics Lab)
2. The slides for this and other talks and articles are available at `personality-project.org/sapa`.
3. The data are available as part of the *psych* package or at Dataverse.
4. The R code is included as an appendix to this talk.

R code

```
#get the current public version of psych from CRAN
install.package(psych)
install.package(psychTools)

## or
 #To get the most recent development release of psych from the
#personality-project.org repository
install.packages("psych",repos="https://personality-project.org/r",
     type="source")
 #Note that you need to restart after installing

library(psych)
sessionInfo() #to show status of R packages
```

```
sessionInfo() #to show status of R packages
R version 4.0.2 (2020-06-22)
Platform: x86_64-apple-darwin17.0 (64-bit)
Running under: macOS Catalina 10.15.7

Matrix products: default
BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
```

do the regressions

```
Random number generation:
 RNG:     Mersenne-Twister
 Normal:  Inversion
 Sample:  Rounding

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] psychTools_2.0.9 psych_2.0.10

loaded via a namespace (and not attached):
[1] compiler_4.0.2  tools_4.0.2     parallel_4.0.2  foreign_0.8-80  tmvnsim_1.0-2
[6] nlme_3.1-148    mnormt_2.0.1    grid_4.0.2      lattice_0.20-41
```

Now score the spi data and do various regressions with it.

R code

```
sc <- scoreItems(spi.keys,spi) # give alpha
mean(sc$alpha[1:5] )#just the big 5
mean(sc$alpha[6:32]) #average alpha for the little 27

R <- cor(sc$scores)
corPlot(R[6:32,1:5],symmetric=FALSE,main=
      "Little 27 and the Big 5 from the SPI")
```

do the regressions

```
ba <- bassAckward(spi[,11:145],c(2,5,27))
     #combine with scores with demographics
sc.demos <-cbind(spi[1:10],sc$scores)
set.seed(42)  #for reproducible results
ss <- sample(1:nrow(sc.demos),nrow(sc.demos)/2)
#derivation multiple Rs
sc.5 <- setCor(y=1:10,x=11:15,, data=sc.demos[ss,], plot=FALSE)
sc.27 <- setCor(y=1:10,x=16:42, data=sc.demos[ss,], plot=FALSE)
sc.135 <- setCor(y=1:10,x=11:145,data=spi[ss,] ,plot=FALSE)

#now cross validate
cv.5 <- crossValidation(sc.5,sc.demos[-ss,])
cv.27 <- crossValidation(sc.27,sc.demos[-ss,])
cv.135 <- crossValidation(sc.135,spi[-ss,])

cross.valid.df <- data.frame(cv5=cv.5$crossV, cv.27=cv.27$crossV,
                  cv135=cv.135$crossV)
cross.valid.df.sorted <- dfOrder(cross.valid.df,1)
#show it
 matPlot(cross.valid.df.sorted[c(1,3,5)],
     main="Cross validation of multiple regression on spi data",
       xlas=3, ylab="R")
```

```
 legend(1,.6,cs(135,27,b5),lty=c(3,2,1),col=c(3,2,1))
#now do a bestScales approach
bs <- bestScales(spi[ss,],criteria=colnames(spi)[1:10], folds=10,
          n.item=20,dictionary=spi.dictionary,cut=.05)
bs.cv <- crossValidation(bs,spi[-ss,])

cross.valid.df.bs <- cbind(cross.valid.df,bs=bs.cv$crossV)
cv.df.bs.sorted <- dfOrder(cross.valid.df.bs,1)
 matPlot(cv.df.bs.sorted[c(1,3,5,7)],
    main="Cross validation of multiple regression on spi data",
        xlas=3, ylab="Cross Validated R")
legend(1,.6,cs(135,27,bestS,b5),lty=c(3,2,4,1),col=c(3,2,4,1))
```

Manhattan plots of the persome: Predict 3 criteria

R code

```
labels <- names(spi.keys)
labels <- abbreviate(labels,minlength=8)
op <- par(mfrow=c(2,3)) #two row by three column display
man <- manhattan(spi,criteria=cs(health,exer,smoke),
     keys=spi.keys,abs=FALSE,labels=labels)
man <- manhattan(spi,criteria=cs(health,exer,smoke),
     keys=spi.keys,abs=FALSE,
```

```
    labels=labels,log.p = TRUE,main="")
op <- par(mfrow=c(1,1)) #put it back to the normal condition
```

Now find the phenotypic and profile correlations

R code

```
Rpheno <- corPlot(spi[1:10],scale=FALSE,upper=FALSE,main="spi crite
R <- cor(spi[,11:145],spi[,1:10],use="pairwise")
R.profile <- corPlot(R,upper=FALSE,scale=FALSE)
corPlot(lowerUpper(Rpheno,R.profile),main='phenotypic and profile c
```

Now, do this for the 126K cases in the bigger sapa data set We get this by going to Condon & Revelle (2015); Condon et al. (2017a,b) and getting the 3 rda files there. We then stitch these three together using rbind to create the full.sapa data

R code

```
sapa <- read.file()  #goes to my directory to find the file
load(sapa) #one extra step required
```

Replicate on a much larger data set.

```
sapa <- char2numeric(sapa)  #makes the fields numeric
criteria <- colnames(sapa)[c(2:10,14:23)]  #choose 19 criteria
spi.items <- selectFromKeys(spi.keys)
options("mc.cores"=8)   #I am using a mac with multiple cores
scores <- scoreIrt.2pl(spi.keys,sapa)  # IRT scoring of the data

big.scores <- cbind(sapa[criteria],scores)

set.seed(42)  #for reproducible results
ss <- sample(1:nrow(big.scores),nrow(big.scores)/2)

#derivation multiple Rs
sc.5 <- setCor(y=criteria,x=20:24, data=big.scores[ss,],
    plot=FALSE)
sc.27 <- setCor(y=criteria,x=25:51, data=big.scores[ss,],
     plot=FALSE)
sc.135 <- setCor(y=criteria, x=spi.items,data=sapa[ss,] ,
   plot=FALSE)

#now cross validate
cv.5 <- crossValidation(sc.5,big.scores[-ss,])
cv.27 <- crossValidation(sc.27,big.scores[-ss,])
cv.135 <- crossValidation(sc.135,sapa[-ss,])
```

Replicate on a much larger data set.

```
cross.valid.df <- data.frame(cv5=cv.5$crossV, cv.27=cv.27$crossV,
       cv135=cv.135$crossV)
cross.valid.df.sorted <- dfOrder(cross.valid.df,1)
#show it
# matPlot(cross.valid.df.sorted[c(2,4,6)],
   main="Cross validation of multiple regression on sapa data",
       xlas=3, ylab="Cross Validated R")
 #legend(1,.6,cs(135,27,b5),lty=c(3,2,1),col=c(3,2,1))
#now do a bestScales approach with all 696 items
#ItemInfo is a dictionary of the sapa items
bs.sapa<- bestScales(sapa[ss,],criteria=criteria, folds=10, n.item=
dictionary=ItemInfo[,1:2],cut=.05)
bs.cv <- crossValidation(bs.sapa,sapa[-ss,])


#combine the best scales
cross.valid.df <- data.frame(cv5=cv.5$crossV, cv.27=cv.27$crossV,
        cv135=cv.135$crossV,cvbs= bs.cv$crossV)
cross.valid.df.sorted <- dfOrder(cross.valid.df,1)
 matPlot(cross.valid.df.sorted[c(2,4,6,8)],
    main="Cross validation of multiple regression on sapa data",xla
       ylab="Cross Validated R",pch=15:18)
```

Replicate on a much larger data set.

```
legend(1,.5,cs(bestS,27,135,b5),lty=c(4,2,3,1),col=c(4,2,3,1),pch=c
#now try profiles
R.big <- cor(sapa[ss,24:719],sapa[ss,criteria],use="pairwise")
R.pheno <- cor(sapa[ss,criteria],use="pairwise")
 R.profile <- cor(R.big)
sapa.pheno.profile <- lowerUpper(R.pheno,R.profile)
corPlot(sapa.pheno.profile,xlas=3,
    main="Phenotypic (lower) and Profile (upper) correlations")
```

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.

Block, J. (1971). *Lives through time*. Berkeley: Bancroft Books.

Block, J. (2002). *Personality as an affect-processing system : toward an integrative theory*. Mahwah, N.J.: L. Erlbaum.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, *3*(3), 296–322.

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

Comrey, A. L. (1995). *Revised manual and handbook of interpretations for the Comrey Personality Scales*. San Diego, CA: Educational and Industrial Testing Service.

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. (2018). The SAPA Personality Inventory: An empirically-derived, hierarchically-organized self-report personality assessment model.

Condon, D. M., Doebler, P., Holling, H., Gühne, D., Rust, J., Stillwell, D., Sun, L., Chan, F., Loe, A., & Revelle, W. (2014). International Cognitive Ability Resource. *https://icar-project.com*.

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. (2015). Selected personality data from the SAPA-Project: 08dec2013 to 26jul2014. *Harvard Dataverse*.

Condon, D. M., Roney, E., & Revelle, W. (2017a). Selected personality data from the sapa-project: 22dec2015 to 07feb2017. [48,350 participant data file and codebook]. *Harvard Dataverse*.

Condon, D. M., Roney, E., & Revelle, W. (2017b). Selected personality data from the sapa-project: 26jul2014 to 22dec2015. [54,855 participant data file and codebook]. *Harvard Dataverse*.

Corr, P. J. (2002). J. A. Gray's reinforcement sensitivity theory: Tests of the joint subsystems hypothesis of anxiety and impulsivity. *Personality and Individual Differences*, *33*(4), 511–532.

Costa, P. T. & McCrae, R. R. (1992). *NEO PI-R professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571–582.

DeYoung, C. G. (2010). Toward a theory of the big five. *Psychological Inquiry*, *21*(1), 26–33.

DeYoung, C. G. (2015). Cybernetic big five theory. *Journal of Research in Personality*, *56*, 35–58.

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. & Takemoto-Chock, N. K. (1981). Factors in the natural language of personality: Re-analysis, comparison, and interpretation of six major studies. *Multivariate behavioral research*, *16*(2), 149–170.

Dworak, E. M. & Revelle, W. (2020). Astrology versus astronomy: Comments on Ashton and Lee. *European Journal of Personality*, *34*, 522–523.

Dworak, E. M., Revelle, W., Doebler, P., & Condon, D. M. (2020). Using the International Cognitive Ability Resource as an open source tool to explore individual differences in cognitive ability. *Personality and Individual Differences*.

Elleman, L. G., McDougald, S., Revelle, W., & Condon, D. (2020). That takes the biscuit: A comparative study of predictive accuracy and parsimony of four statistical learning techniques in personality data, with data missingness conditions. *European Journal of Psychological Assessment (in press)*.

Eysenck, H. J. (1967). *The biological basis of personality*. Springfield: Thomas.

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.

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. & Saucier, G. (2016). The Eugene-Springfield Community Sample:
Information available from the research participants. Technical Report 56-1, Oregon Research Institute, Eugene, Oregon.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, *37*(6), 504 – 528.

Gough, H. G. (1957). Manual for the California Psychological Inventory.

Gray, J. A. (1991). The neuropsychology of temperament. In J. Strelau & A. Angleitner (Eds.), *Explorations in temperament: International perspectives on theory and measurement* (pp. 105–128). New York, NY: Plenum Press.

Hathaway, S. & McKinley, J. (1943). Manual for administering and scoring the MMPI.

Hogan, R. (1982). A socioanalytic theory of personality. In *Nebraska Symposium on Motivation* (pp. 55–89). University of Nebraska Press.

Hogan, R. (2007). *Personality and the fate of organizations*. ix, 167 pp. Mahwah, NJ: Lawrence Erlbaum Associates Publishers Lawrence Erlbaum Associates Publishers.

Hogan, R. & Blickle, G. (2018). Socioanalytic theory: Basic concepts, supporting evidence and practical implications.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory-Versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institue of Personality and Social Research.

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.

Kluckhohn, C. & Murray, H. A. (1948). *Personality in nature, society, and culture*. New York: A. A. Knopf.

Konstabel, K., Lönnqvist, J.-E., Leikas, S., Velàzquez, R. G., H, H. Q., Verkasalo, M., , & et al. (2017). Measuring single constructs by single items: Constructing an even shorter version of the "short five" personality inventory. *PLoS ONE*, *12*(8), e0182714.

Lord, F. M. (1955a). Estimating test reliability. *Educational and Psychological Measurement*, *15*, 325–336.

Lord, F. M. (1955b). Sampling fluctuations resulting from the sampling of test items. *Psychometrika*, *20*(1), 1–22.

Lord, F. M. (1977). Some item analysis and test theory for a system of computer-assisted test construction for individualized instruction. *Applied Psychological Measurement*, *1*(3), 447–455.

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.

Ozer, D. J. & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, *57*, 401–421.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rammstedt, B. & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in English and German. *Journal of Research in Personality*, *41*(1), 203 – 212.

Revelle, W. (1995). Personality processes. *Annual Review of Psychology*, *46*, 295–328.

Revelle, W. (2019). Personalty research: an open and shared science. Invited presentation to the Department of Psychology, University of Zurich.

Revelle, W. (2020). *psych: Procedures for Personality and Psychological Research*. https://CRAN.r-project.org/package=psych: Northwestern University, Evanston. R package version 2.0.8.

Revelle, W. & Condon, D. M. (2017). Using MMCAR to explore the structure of personality and ability. In *Annual Meeting of the International Psychometric Society*, Zurich.

Revelle, W. & Condon, D. M. (2019). Reliability from alpha to omega: A tutorial. *Psychological Assessment.*, *31*(12), 1395–1411.

Revelle, W., Condon, D. M., & Heintz, S. (2018). Assessing personality with massively missing completely at random data:

An information theoretic approach. In *Experts meeting on personality measurement*, Edinburgh, September 6-8, 2018.

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* (2nd ed.). chapter 37, (pp. 578–595). Sage Publications, Inc.

Revelle, W., Dworak, E. M., & Condon, D. M. (2020a). Cognitive ability in everyday life: the utility of open source measures. *Current Directions in Psychological Science*, *29*(4), 358–363.

Revelle, W., Dworak, E. M., & Condon, D. M. (2020b). Exploring the persome: The power of the item in understanding personality structure. *Personality and Individual Differences*.

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.

Sackett, P. R. & Kuncel, N. R. (2018). Eight myths about standardized admissions testing. In B. W. Jack Buckley, Lynn Letukas (Ed.), *Measuring Success: Testing, Grades, and the Future of College Admissions* chapter 1, (pp. 13–38). Baltimore, MD: Johns Hopkins University Press.

Saucier, G. (1994). Mini-markers: A brief version of goldberg's unipolar big-five markers. *Journal of Personality Assessment*, *63*(3), 506–516.

Soto, C. J. & John, O. P. (2017). The next big five inventory (bfi-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social psychology*, *113*(1), 117–143.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology*, *15*(2), 201–292.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *3*(3), 271–295.

Tryon, R. (1957). Communality of a variable: Formulation by cluster analysis. *Psychometrika*, *22*(3), 241–260.

Tryon, R. (1959). Domain sampling formulation of cluster and factor analysis. *Psychometrika*, *24*(2), 113–135.

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).

Waller, N. G. (2008). Fungible weights in multiple regression. *Psychometrika*, *73*(4), 691–703.