Cognadev Technical Report Series

14

18th October, 2018

Equifinal Profiling

I was asked this question recently by an executive responsible for hiring in a large corporate:

"We observe too often that people with seemingly disparate profiles can excel in the same role. What type of analyses can someone do with a big dataset of predictors and criteria to determine whether multiple "profiles" can predict success? It seems that traditional model approaches can't do this as they just create a single 'average' profile or solution."

Cognadev UK Ltd., 2 Jardine House, Harrovian Business Village, Harrow, Middlesex, HA1 3EX. Registered UK # 05562662

Definition

Equifinality is a term from general systems theory, first proposed by Ludwig von Bertalanffy in 1968. It is defined as the property of a system to reach similar outcomes from different starting points through different processes. Put another way, a single outcome may be the result of different 'compensatory' mixes of input attributes. The particular combinations of these attributes, and their relative magnitudes, are causal for a single outcome. For example, if we consider a single outcome such as 'successful employee hire', where the outcome is adjudged 9 months after the hiring decision, equifinality may be represented as:

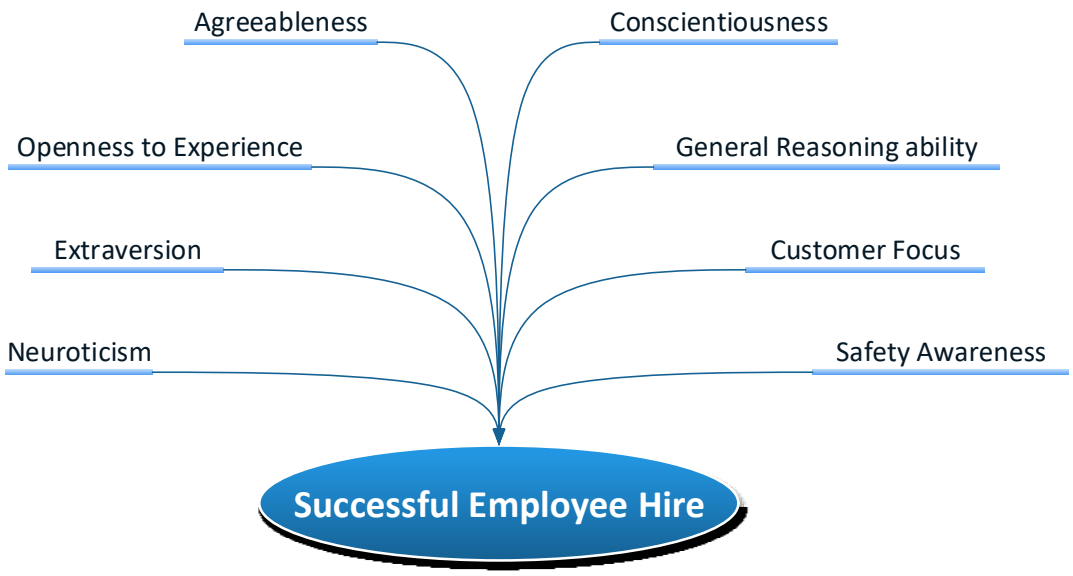

Figure 1: multiple hypothesised causes of a successful employee hire

Up to 8 attributes are hypothesised as contributing to the outcome. A 'closed system' view would state that only one 'profile' of attribute magnitudes is causal for the outcome. An 'open system' *equifinality* viewpoint would state that more than one profile of attribute magnitudes might be causal for the outcome.

The Article Dataset

Within this article, I'm going to use a small dataset consisting of 12 incumbent employees selected from a larger group, who were assessed using a standard Big five personality questionnaire. The N,E,O,A,C scores range between 0 and 10.The aim of HR was to develop a profile for adjudged successful employees, which could be used in future to help screen-out new hires who did not match the profile according to a profile match statistic. This is a typical post-hoc profile construction scenario, where tests are administered to incumbent employees in order to develop a profile against which new candidates for the job role might be compared.

Table 1: The equifinal profiling dataset used throughout this article

	Equifinal Profiling Dataset					
	1 N	2 E	3 O	4 A	5 C	6 Job Performance
1	2	2	7	7	6	Good
2	2	7	8	6	6	Good
3	2	4	7	9	7	Good
4	3	2	7	7	6	Poor
5	3	7	8	5	7	Poor
6	3	4	7	5	6	Poor
7	6	8	6	3	6	Good
8	6	5	8	2	6	Good
9	6	2	6	3	8	Good
10	4	2	8	7	6	Poor
11	5	7	8	5	7	Poor
12	6	4	9	3	7	Poor

Key

N = Neuroticism/Anxiety

E = Extraversion

O = Openness to Experience

A = Agreeableness

C = Conscientiousness

Conventional Profiling

1 One of the common approaches to profile construction by HR and I/O consultants is to construct the "profile for success". What normally happens is that a sample of the top-performing incumbent employees are selected, their scores on one or more assessments acquired, and a profile constructed using the mean scores of the 'star performers', against which new candidate-employee test-scores are compared. In this situation, the HR analyst would only see the 'Good' cases in the dataset on page 3.

	Equifinal Profiling Dataset - Top Performer subset					
	1 N	2 E	3 O	4 A	5 C	6 Job Performance
1	2	2	7	7	6	Good
2	2	7	8	6	6	Good
3	2	4	7	9	7	Good
4	6	8	6	3	6	Good
5	6	5	8	2	6	Good
6	6	2	6	3	8	Good

Table 2: The "good performer' subset of the equifinal profiling dataset used throughout this article

To construct the 'top performer profile', I now use the mean scores on each attribute:

	Job Performance=Good Descriptive Statistics (multiple success profiles.sta)	
Variable	Valid N	Mean
N	6	4.0
E	6	4.7
O	6	7.0
A	6	5.0
C	6	6.5

Table 3: The mean scores for the "good performer' employee subset

The graph of the profile is:

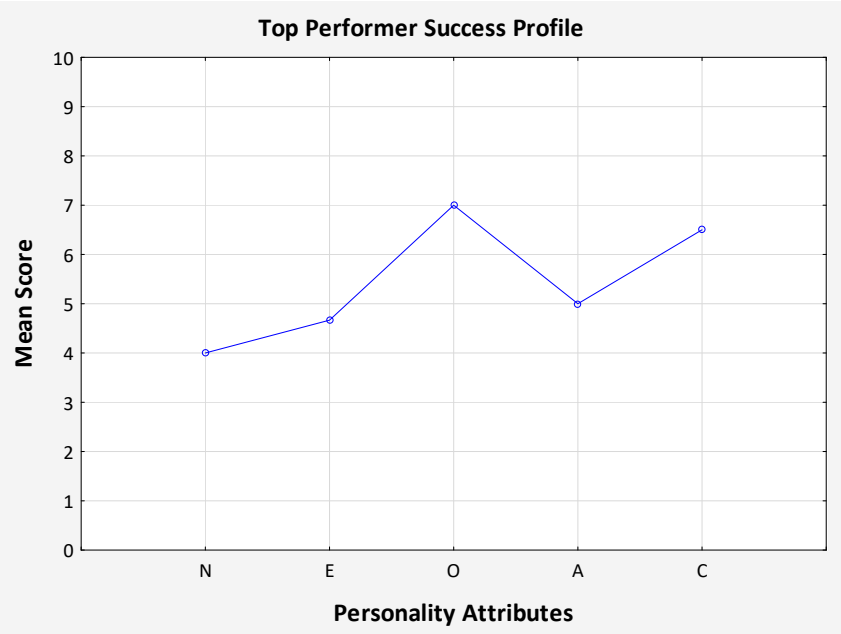

Figure 2: The success profile for the "good performer" employee subset

Two potential uses of this profile might now follow:

1. Mean scores on O and C look elevated compared to the others, so we select candidates who score as high or higher on O and C (and these are now spoken of as our 'success factors').

2. The profile as a whole is used as a target vector against which every new candidate profile is compared; using correlation, a profile similarity coefficient, or some other index of 'closeness to target'.

❷ The problem with ❶ above is that without investigating how those who are not adjudged 'top performers' score on the assessment, it is perfectly possible that these 'non-selected' individuals score exactly the same or even more in the 'success' direction as those who constitute the success profile. So, here, we produce a contrast profile, by selecting a sample of poor performers and produce the 'poor/failure' profile. Then we note those attributes which discriminate between poor and good performers, and use these as our 'attributes' which will predict success.

	Equifinal Profiling Dataset - Poor Performer subset					
	1 N	2 E	3 O	4 A	5 C	6 Job Performance
1	3	2	7	7	6	Poor
2	3	7	8	5	7	Poor
3	3	4	7	5	6	Poor
4	4	2	8	7	6	Poor
5	5	7	8	5	7	Poor
6	6	4	9	3	7	Poor

Table 4: The "poor performer' subset of the equifinal profiling dataset used throughout this article

To construct the 'poor performer profile', I now use the mean scores on each attribute:

	Job Performance=Poor Descriptive Statistics (multiple success profiles.sta)	
Variable	Valid N	Mean
N	6	4.0
E	6	4.3
O	6	7.8
A	6	5.3
C	6	6.5

Table 5: The mean scores for the "good performer' employee subset

The good and poor performer profiles might be presented as spline plots ..

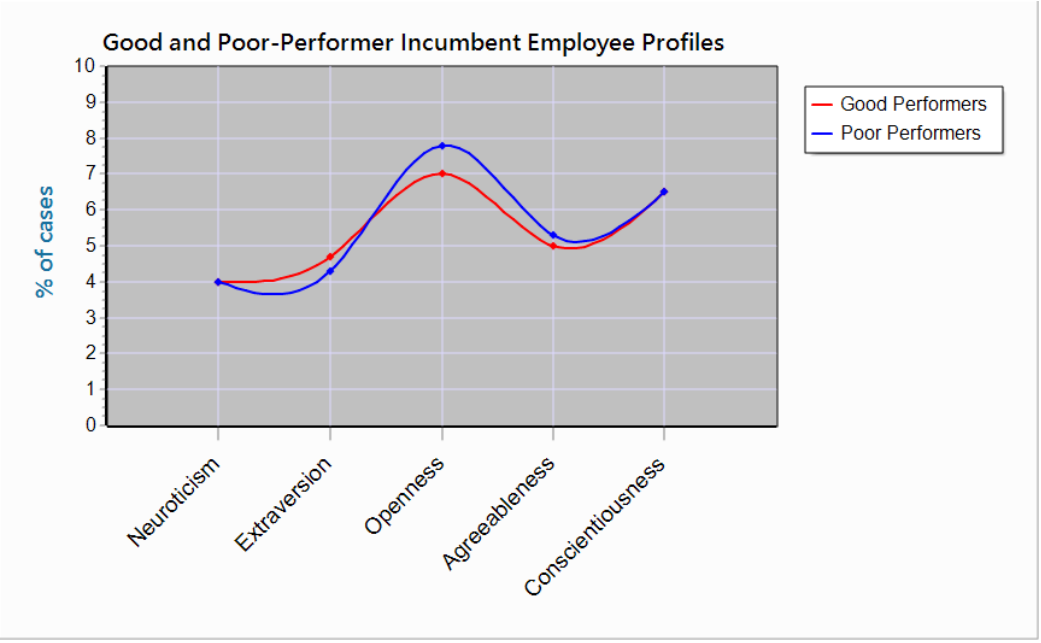

Figure 3: The profiles for "good" and "poor" performers

What you notice immediately from Figure 3 is that there is almost no difference between the profiles of the good and poor performers.

Whereas if we had just constructed a "success" profile from our top performers without looking at the poor performers (or those not classed as "top performers"), we might have assumed that selecting candidates who score as high or higher on O and C would help us select likely top performers.

However, having computed the profile from the poor performers, we can see that their mean is actually higher on O than within the 'success profile' and exactly the same for C.

If we had compared all our cases against the success profile in Figure 2, using a variety of profile comparison indices, we would see:

Case #	Pearson r	Cattell r_p	Gower	KSD-S_{sharp}	ICC_1	ICC_2	ICC_3	Criterion
1	.75	.40	.86	.64	.62	.61	.59	Good
2	.71	.44	.86	.70	.65	.64	.60	Good
3	.61	.32	.86	.68	.53	.51	.46	Good
4	.71	.42	.88	.71	.64	.63	.59	Poor
5	.81	.53	.90	.80	.73	.73	.73	Poor
6	.98	.82	.96	.94	.92	.92	.95	Poor
7	-.02	.02	.82	.58	.08	-.02	-.02	Good
8	.51	.32	.86	.69	.52	.49	.44	Good
9	.52	.28	.82	.55	.49	.46	.42	Good
10	.71	.44	.88	.71	.65	.63	.58	Poor
11	.79	.33	.90	.80	.60	.64	.78	Poor
12	.71	.43	.86	.67	.64	.64	.59	Poor

Table 6: Profile 'match' indices: comparing 12 case profiles to the 'success' Target profile

Notes:

- The Cattell r_p index used raw data, and a chi-square expected value based upon the raw Sds's for each profile.
- The Kernel Smoothed Distance- similarity (KSD-s) used a sharpness SD of 1.6667.
- $ICC_{1, 2, and 3}$ are intraclass correlations computed using Fleiss's model 1, 2, and 3 formulae.
- The Pearson r, r_p, and ICC's can vary between -1.0 and +1.0 *(negative values are very rare for ICCs)*, where 0 = no relationship/association and 1.0 = perfect monotonic association. Contrary to the received wisdom, no ICC model measures agreement; see the example on page 12 of http://www.pbarrett.net/presentations/Interrater_Reliability_Barrett_Hogan_Hogan_ISSID_2009.pdf
- Relative to the maximum possible absolute (unsigned) discrepancy between the two pairs of observations, the Gower discrepancy coefficient indicates the % average absolute discrepancy between all pairs of observations. When expressed as a similarity coefficient (by subtracting it from 1), a Gower of 0.90 for example indicates that relative to the maximum possible absolute (unsigned) discrepancy between them, the observations agree, on average, to within 90% of each other's values. The Gower varies between 0 (maximal dissimilarity) to 1.0 (magnitude identity).
- The KSD-s is based upon a very simple idea that a distance function should be shaped in such a way that if the simple arithmetic unsigned difference between a person's attribute value and a target value is

computed to be within a certain range, then the computed distance should reflect a very small distance, almost regardless of the actual distance. But, as that distance grows larger, then the computed distance should be accelerated in magnitude. In short, a non-linear "inertial" effect is created – translated into a distance metric. The coefficient itself is scaled as a measure of similarity, varying between 0 (maximal dissimilarity) to 1.0 (magnitude identity).

Formulae for all coefficients are provided in an Appendix to this article.

From Table 6, if we were to use a reasonable threshold for each coefficient:

- Pearson r, Cattell r_p, $ICC_{1, 2, \& 3}$: select cases > 0.70 as 'good enough' match to the success profile
- Gower: select as 'good enough' > 0.90
- KSD-s: probably select > 0.80 as 'excellent match'

Case #	Pearson r	Cattell r_p	Gower	KSD-S_{sharp}	ICC_1	ICC_2	ICC_3	Criterion
1	.75							Good
2	.71							Good
3								Good
4	.71							Poor
5	.81		.90	.80	.73	.73	.73	Poor
6	.98	.82	.96	.94	.92	.92	.95	Poor
7								Good
8								Good
9								Good
10	.71							Poor
11	.79		.90	.80			.78	Poor
12	.71							Poor

Table 7: Selected cases for 'good enough' match to the success-profile.

With a base-rate of 0.50, the predictive efficiency (overall classification accuracy) of each coefficient is:

Coefficient	PE	PPP	NPP	Relative Improvement over Chance (RIOC)
Pearson r	17%	25%	0%	-1.33
Cattell r_p	42%	0%	45%	-833333.33
Gower	25%	0%	33%	-7500000.0
KSD-s	25%	0%	33%	-7500000.0
ICC_1	33%	0%	40%	-3333333.33
ICC_2	33%	0%	40%	-3333333.33
ICC_3	25%	0%	33%	-7500000.0

Table 8: The relevant actuarial statistics indicative of profile match success, for each profile matching coefficient.

Notes:

PE: Predictive Efficiency (overall predictive accuracy)

PPP: Positive Power to Predict (the accuracy of predicting Good/Successful employees)

NPP: Negative Power to Predict (the accuracy of predicting Poor/Unsuccessful employees)

RIOC: Loeber and Dishion's statistic[1] indexes the improvement of prediction over chance, relative to the Base Rate, using your test. This is an extremely valuable statistic that gives a clear indication of just how good your profiling is in terms of predictive accuracy. Whereas the Improvement Over Chance (IOC) coefficient indexes the basic improvement over chance, the IOC index is sensitive to both Base Rate and Selection Ratio. By expressing the IOC relative to the maximum possible accuracy (the Base Rate) given the lowest possible accuracy (chance), the **RIOC** provides a universal measure of effect which is much less dependent upon sample characteristics than the IOC.

Put another way, the RIOC expresses the improvement over chance (IOC) as a function of the difference between the random correct (RC) and maximum correct (MC) values in a given study. Thus, the percent improvement over chance in a given study always falls between the random correct value and the maximum correct value. An RIOC of 0.0 would indicate that your selections are no better than chance (or random allocation). An RIOC of > 0.0 indicates increasing improvement over chance levels of prediction. The coefficient has no upper limit. Negative RIOCs indicate the selection procedure is actually worse than tossing a coin to select each candidate.

Quite clearly, none of these profile matching coefficients seem able to cope with multiple profiles which are associated for a single outcome (the equifinal principle). However, using a more correct methodology, it is possible to be 100% accurate in predicting candidates who will be successful (based upon the assumption that the personality scores are the sole cause for success/failure (ratings of 'Good' and 'Poor').

[1] Loeber, R. & Dishion, T. (1983) Early Predictors of male delinquency: a review. *Psychological Bulletin*, 94, 68-99
Mossman, D. (1994) Assessing Predictions of Violence: being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, 62, 4, 783-792.

Equifinal Profiling and Decision Trees

Let's refresh ourselves with the data we are working with (from Table 1 above):

	Equifinal Profiling Dataset					
	1 N	2 E	3 O	4 A	5 C	6 Job Performance
1	2	2	7	7	6	Good
2	2	7	8	6	6	Good
3	2	4	7	9	7	Good
4	3	2	7	7	6	Poor
5	3	7	8	5	7	Poor
6	3	4	7	5	6	Poor
7	6	8	6	3	6	Good
8	6	5	8	2	6	Good
9	6	2	6	3	8	Good
10	4	2	8	7	6	Poor
11	5	7	8	5	7	Poor
12	6	4	9	3	7	Poor

And what the data (as profiles) look like compared to our 'success' Target profile:

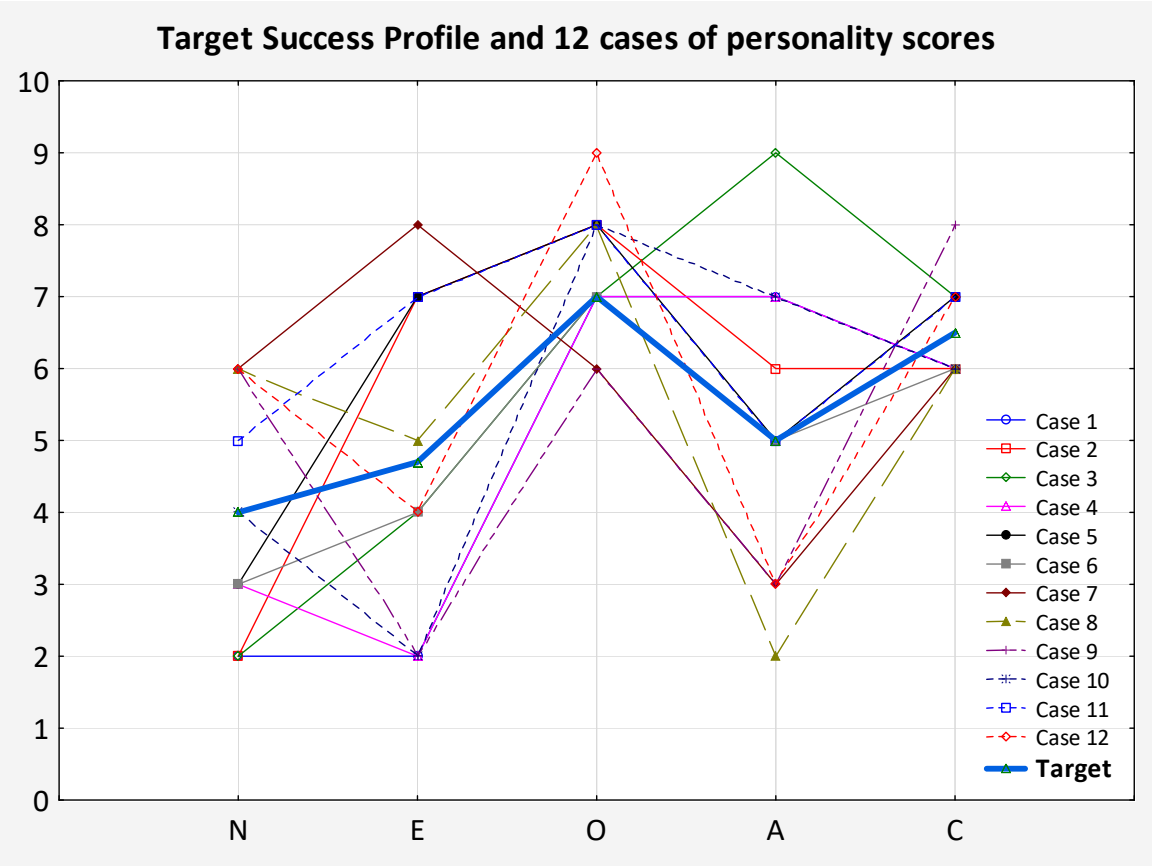

Figure 4: The 12 cases of profile data, and the target 'success' profile.

Looking at this plot, it is no wonder we see indifferent profile matching from whatever may be our favoured coefficient.

Now let's use a new technique to attempt to predict both the cases whose personality scores are predictive of 'success', and those predictive of 'failure' (where *Good* ratings are adjudged 'success', and *Poor* ratings as 'failure').

A decision tree is a structure built from a series of decisions that aim to maximize the classification accuracy of two or more outcome classes, levels, or measures. The analogy with the form of a tree is what gives the structure its name. The concept and major algorithms for a constructing a decision tree were introduced by Breiman, Friedman, Olshen, and Stone (1984) with the Classification and Regression Tree algorithm (CART), and Quinlan (1986, 1993) with the ID3 and C.4.5 classifier algorithms.

Here I'm using Breiman's algorithm implemented in Statistica 12, with a simple 'decision-split' algorithm option rather than forming 'regression trees'. More details about decision trees can be found in Barrett (2005)[2] and Strobl, Malley, & Tutz (2009).[3]

The decision tree computed on the 12-case data shows 100% classification accuracy. Clearly, with artificially constructed data designed to show a principle, rather than with real data, such accuracy must be tempered with caution!

Ordinarily, fitting a decision tree has to be cautious, as the capacity for over-fitting/capitalization on sample-data-specific features is substantive; what might be called 'over-learning'. Usually, intensive cross-validation procedures (e.g. v-fold or holdout-sample calibrations) are undertaken before a final solution is proposed (see Koul, 2018)[4]. Anyway, the point of this analysis is simply to show what a technique like this can do, if certain patterns exist within datasets which are predictive of specific outcomes.

Figure 5 on the next page shows the decision tree for the 12-case dataset.

[2] In André Beauducel, Bernhard Biehl, Michael Bosnjak, Wolfgang Conrad, Gisela Schönberger, and Dietrich Wagener (Eds.) *Multivariate Research Strategies: a Festschrift for Werner Wittman*. Chapter 4, pp 63-118. Aachen: Shaker-Verlag. (http://www.pbarrett.net/publications/Person_Target_Profiling_Barrett_2005.pdf).

[3] Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14, 4, 323-348.

[4] Koul, A. (2018). Cross-validation approaches for replicability in psychology. *Frontiers in Psychology: Quantitative Psychology and Measurement* (https://doi.org/10.3389/fpsyg.2018.01117), 9, 1117, 1-4.

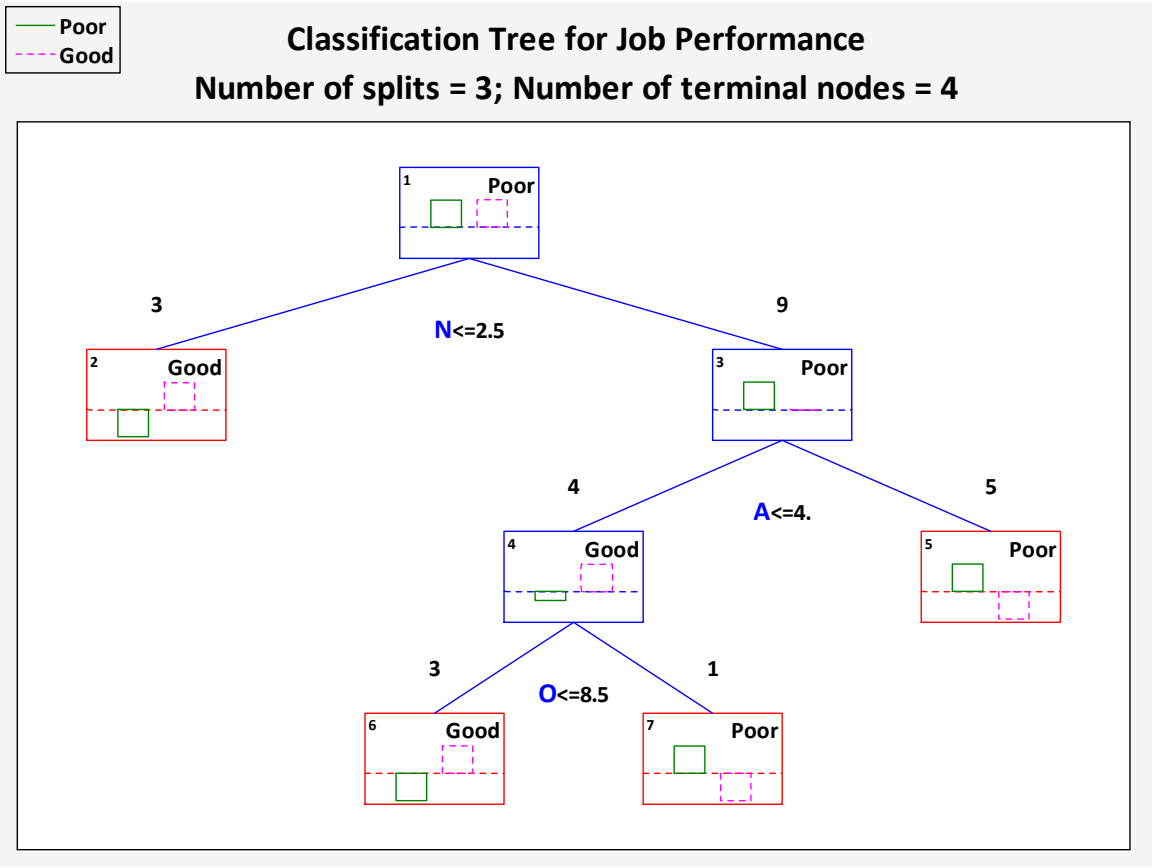

Figure 5: The 100% accurate decision-tree for the 12-case dataset. The scores on only three personality factors are required to attain 100% accuracy of performance prediction; N (Neuroticism), A (Agreeableness), and O (Openness to Experience).

The 'terminal nodes' of the tree define 'profiles'. But these are now a function of sequential *univariate splits*, not of mean scores or profile match coefficients based upon analysis of vectors of scores.

Profile ①

We just use the **N** scale as a single-scale 'profile', and select those cases with a score of 2 or less on N. The results show we would correctly identify 50% of our 'successful' employees, and reject all 'Poor' Employees correctly. Overall predictive accuracy would be 75%.

Rule	PE	PPP	NPP	RIOC
N < 3	75%	100%	67%	infinity

Profile ②

If we now introduce scale **A** after extracting successful cases using profile ①, the results show that by applying the dual-component sequential rule, we would correctly identify 100% of our 'successful' employees, and reject five 'Poor' Employees correctly, but mistakenly include a single poor employee in our 'success' candidate list (a false-positive). Overall predictive accuracy would now be 92%.

Rule	PE	PPP	NPP	(RIOC)
(N < 3) then (A <= 4)	92%	86%	100%	2.92

Profile ③

If we now introduce scale **O** after extracting successful cases using profile ② , the results show that by applying the tri-component *sequential* rule, we would correctly identify 100% of our 'successful' employees, and correctly reject 100% of our 'Poor' rated Employees correctly. Overall predictive accuracy would now be 100%.

Rule	PE	PPP	NPP	(RIOC)
(N < 3) then (A <= 4) then (O < 9)	100%	100%	100%	infinity

I know; in reality, these are not 'profiles' at all … rather they are decision rules applied in sequence which optimize the prediction of two classes of outcome – success and failures.

If we were using more predictors, and had many more cases, we might well have formed 'profiles' at each stage, formed from splits computed from optimal regression models of predictors rather than single attributes (as did here). Remember, this whitepaper is all about 'principle'; it's not a recipe for 'how to'!

While we could all do the above working by-hand and by-eye, consider the problem when you have hundreds of cases to analyse, and perhaps 20 or more of profile variables, some not scores at all but categories and/or ordered classes. CART is a perfect exemplar of machine learning – we have a machine to do what we would do 'by-hand and by-eye' if we could only cope!

And I'm sure there is still the nagging feeling that if we could hunt through data selectively, we might be able to find cases who share complex mixes of attribute-magnitudes in common, forming homogenous groups which differ from other groups, but where those groups are also predictive of the same outcome. Welcome to inductive profiling.

Would Inductive Profiling work?

Inductive profiling is used where the goal is to find 'homogenous groups' of individuals within a dataset, then explore what that homogeneity means viz a viz outcome criteria. More familiar methods to psychologists will be cluster analysis, multivariate profile analysis, and Q-factor analysis. However, all these methods are data-model bound, using statistical distribution summary parameters and concepts, and as with Q-factor analysis, working with transformed/normalized data rather than with the actual observation magnitudes.

However, modern methods no longer rely upon such constraints for their implementation, where control over both the algorithms and what should constitute the threshold for 'homogenous' can be varied iteratively and in an exploratory fashion. Kohonen neural networks do this, as well as more simple 'hunting' algorithms. Here I'm using a Gower coefficient to show how it can be implemented very simply, without moving into a more refined procedure with a set of 'if-then' production rules designed to 'sharpen up' what we wish to define as 'homogenous'.

Usually, inductive profiling is 'criterion free'; its goal being to identify groups of cases which share something in common. Then we investigate other relevant characteristics that is associated with the groups (if any).

Initially, I'll compute the Gower agreement matrix between all cases (every case compared to every other case, across the five Big Five profile attributes N, E, O, A, and C).

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10	Case 11	Case 12
Case 1	1.00	0.86	0.90	0.98	0.80	0.90	0.70	0.74	0.78	0.94	0.76	0.74
Case 2		1.00	0.84	0.84	0.94	0.88	0.80	0.80	0.68	0.84	0.90	0.76
Case 3			1.00	0.88	0.82	0.88	0.68	0.72	0.72	0.84	0.78	0.76
Case 4				1.00	0.82	0.92	0.72	0.76	0.80	0.96	0.78	0.76
Case 5					1.00	0.90	0.82	0.82	0.74	0.82	0.96	0.82
Case 6						1.00	0.80	0.84	0.80	0.88	0.86	0.84
Case 7							1.00	0.88	0.84	0.72	0.86	0.84
Case 8								1.00	0.84	0.80	0.86	0.92
Case 9									1.00	0.80	0.78	0.88
Case 10										1.00	0.82	0.80
Case 11											1.00	0.86
Case 12												1.00

Table 9: Gower indices comparing each case with every other case, using a Big Five N, E, O, A, and C score profile.
Note: I've highlighted all coefficients >= 0.90 (i.e. relative to the maximum possible absolute (unsigned) discrepancy between them, the profiles agree, on average, to within 90% of each other's constituent values).

Because this is a relatively small matrix, I could form homogenous clusters 'by eye', looking at coefficients >= .90 between cases in order to identify homogenous groups. Another way is to use non-metric multidimensional scaling, producing a 2-dimensional case-plot from the input similarities.

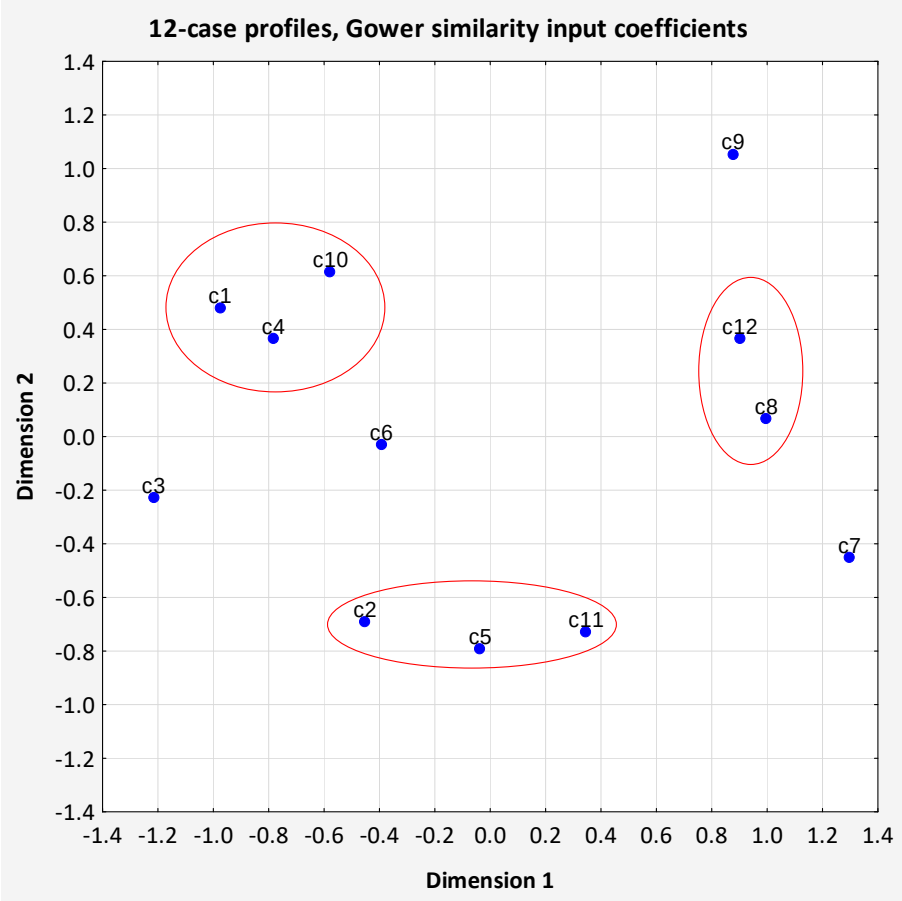

Figure 6: Non-metric MDS solution for 12-case profile comparison similarity matrix

Note: I've subjectively highlighted cases which seem to be clustered together. For real-world profiling, some form of nearest-neighbour analysis would be undertaken on the coordinate values to objectively identify profile clusters.

Not surprisingly, the clusters conform to the largest highlighted Gower coefficients in Table 9:

Case	Equifinal Profiling Dataset- Homogenous Group 1					
	1 N	2 E	3 O	4 A	5 C	6 Job Performance
1	2	2	7	7	6	Good
4	3	2	7	7	6	Poor
10	4	2	8	7	6	Poor
	Equifinal Profiling Dataset- Homogenous Group 2					
	1 N	2 E	3 O	4 A	5 C	6 Job Performance
2	2	7	8	6	6	Good
5	3	7	8	5	7	Poor
11	5	7	8	5	7	Poor
	Equifinal Profiling Dataset- Homogenous Group 3					
	1 N	2 E	3 O	4 A	5 C	6 Job Performance
8	6	5	8	2	6	Good
12	6	4	9	3	7	Poor

Table 10: Three Homogenous-Score Groups identified from the non-metric MDS solution in Figure 6.

And four cases were not sufficiently homogenous with any group – they remain 'independent' cases.

	Equifinal Profiling Dataset- Independent cases					
	1 N	2 E	3 O	4 A	5 C	6 Job Performance
3	2	4	7	9	7	Good
6	3	4	7	5	6	Poor
7	6	8	6	3	6	Good
9	6	2	6	3	8	Good

Table 11: Four independent case profiles, insufficiently close to any other profile.

The problem of course is that although we identified some homogenous groups, the Job Performance criterion categorization is not associated with them. So, while inductive profiling may be useful in other situations, here it is a complete failure.

Bottom line

A decision-tree approach to profiling is fundamentally different from other more conventional forms.

It doesn't reveal multiple profiles in the conventional sense, but sequential attribute 'threshold-split' rules using an optimal subset of profile attributes. It is a particular rule-set which defines the profile, but the rules are intrinsically non-linear in that multiple splits may be made on the same attribute, relative to the splits on other attributes.

So, while not quite the kind of answer to the original question we might have expected, we do have a method which tries to account for the 'patterning' of attribute magnitudes within data rather than just working with a single linear profile vector and unidirectional magnitudes on profile attributes.

Appendix A: The profile coefficient equations

A.1.1 Intraclass Model 1

Each profile attribute score for each attribute is assumed independent of each other, including within each profile. It's not the correct model at all for these kinds of profile data. The ANOVA formula is:

$$r_{icc1}^2 = \frac{MS_p - WMS}{MS_p + (n_r - 1) * WMS}$$

where MS_p = Between profiles mean square

n_r = number of cases x number of profile attributes

n_a = number of attributes

MS_{res} = Residual mean square

WMS = Within attributes mean square

MS_a = Between profile attributes mean square

with

$$WMS = \frac{\left[\left(MS_a * (n_r - 1)\right) + \left(MS_{res} * (n_p - 1) * (n_r - 1)\right)\right]}{n_a * (n_r - 1)}$$

A.1.2 Intraclass Model 2

Every profile is considered a random sample from some population of profiles composed of the five attributes. The ANOVA formula is:

$$r_{icc2} = \frac{MS_p - MS_{res}}{MS_p + (n_p - 1) * MS_{res} + \left(\frac{n_p * \left(MS_a - MS_{res}\right)}{n_a}\right)}$$

where MS_p = Between profiles mean square

MS_{res} = Residual (interaction) mean square

MS_a = Between profile attributes mean square

n_p = The number of profiles

n_a = The number of attributes

A.1.3 Intraclass Model 3

Every profile is considered a unique profile. That is, there is no assumption that profiles have been sampled from some hypothetical population. The ANOVA formula is:

$$r_{icc3} = \frac{MS_p - MS_{res}}{MS_p + (n_p - 1) * MS_{res}}$$

where MS_p = Between Profiles mean square

MS_{res} = Residual (interaction) mean square

n_p = The number of profiles

The Intraclass coefficients vary between 0 and +1 (where 0 = no relationship/reliability at all). They can under certain circumstances assume negative values but this is very rare with anything but random rating data.

A.1.4 The Gower similarity index

Relative to the maximum possible absolute (*unsigned*) discrepancy between the two pairs of observations, the gower *discrepancy* coefficient indicates the % average absolute discrepancy between all pairs of observations. When expressed as a similarity coefficient (by subtracting it from 1), it indicates the % average similarity between all pairs of observations.

So, a Gower similarity coefficient of say 0.90 indicates that relative to the maximum possible absolute (*unsigned*) discrepancy between them, the observations agree, on average, to within 90% of each other's values.

If you change the value of that maximum possible discrepancy, then the Gower coefficient will change to reflect this, as the discrepancies between pairs of observations are divided (scaled) by that maximum possible discrepancy value. E.g. if two observations differ by 5, and the measurement range of each observation is 10, then the relative discrepancy is 0.5. However, if the measurement range for each observation was say 100, then the relative discrepancy would be just 0.1.

But that's the whole point of the Gower, it tells you how discrepant (or similar) observations are, RELATIVE to how discrepant they could have been. A 5-point difference in a 10-point maximum measurement range is not very good. A 5-point difference between observations within a 100-point measurement range is pretty accurate.

One final point, the minimum and maximum possible magnitudes of attribute values for a Gower coefficient should be all be equal, not for mathematical reasons *(as the formula below takes into account the differing range for each attribute)*, but for interpretation reasons. When all attributes are in a common metric, it is straightforward to relate the index magnitude to the actual average discrepancy across all attributes, because now all attributes are 'measured' within the same range.

Where attributes are initially in a different metric, I use common-metric-rescaling; which is a method for linearly rescaling attributes that avoids the non-linearity introduced by a standardizing transformation. Details of how to implement this rescaling is provided in pages 40-46 of the Gower v.1.1 program manual which is available for download from: www.pbarrett.net/Gower/Gower.html. All attributes in this whitepaper were measured over the same range of 0-10, so no rescaling was required.

$$Gower_{similarity} = 1 - \left[\frac{\sum_{i=1}^{n} \left(\frac{\left| profile_{1i} - profile_{2i} \right|}{range_i} \right)}{n} \right]$$

$n =$ the number of cases

$range_i =$ the range of the rating attribute i $(maximum - minimum\ value)$

$profile_{1i}$ = the attribute value i of n for profile1

$profile_{2i}$ = the attribute value i of n for profile 2

The Gower coefficient computes a scaled similarity coefficient, utilizing scaled discrepancies. It varies between 0 and +1, where 0 indicates maximum possible dissimilarity, and +1 is equal to magnitude-identity between the two vectors being compared.

Gower, J. C., 1971. A general coefficient of similarity and some of its properties. *Biometrics* 23:623-637.

A.1.5 Cattell's Profile Similarity coefficient r_p

The profile similarity coefficient (Cattell, 1969), r_p, or what has also been called the *pattern similarity coefficient* (Cattell, Coulter, and Tsuijoka, 1966; Cattell, Eber, and Tatsuoka, 1970; Cattell, 1978) was first introduced by Cattell in 1949.

It was designed by Cattell (taken from: Cattell, Coulter, and Tsuijoka, 1966, p.296) to:

- take into account the metric and number of dimensions comprising the profiles to be compared.
- compare the coefficient with the magnitude to be expected by chance.
- provide a convenient function which behaves (e.g. as regards distribution) in essentially the same general way as a Pearson r, varying from +1.0 indicating complete agreement between profiles to 0 for no relation, and -1.0 for complete inverse relation.

The meaning of an r_p of +1.0 is that two persons or patterns have exactly the same profiles and fall on the same point in multi-dimensional space.

A value of 0 indicates that they fall as far apart as would be expected for any two points taken at random.

A value of -1.0 means that they are at opposite ends of the distribution.

Since the ends of a distribution are ill-defined and asymptotic, the value -1.0 is in actual practice approached but never quite reached, and there is in consequence a small asymmetry (positive skewing) in the distribution of r_p about its median value of 0.

The formula for comparing two individual profiles (using raw scale scores) is:

$$r_p = \frac{E_k - \sum_{i=1}^{k} \left(profile_{1i} - profile_{2i} \right)^2}{E_k + \sum_{i=1}^{k} \left(profile_{1i} - profile_{2i} \right)^2}$$

where

$E_k = \chi^2_{50} \cdot \left(s_1^2 + s_2^2 \right)$

$\chi^2_{50} =$ the inverse cumulative probability Chi-Square distribution value for p=.5 and df = k

$k =$ number of profile attributes

$s_1^2 =$ the variance for profile 1

$s_1^2 =$ the variance for profile 2

A worked example:

t = profile 1
c = profile 2

$$t = \begin{bmatrix} 4 \\ 4.7 \\ 7 \\ 5 \\ 6.5 \end{bmatrix} \quad c = \begin{bmatrix} 3 \\ 4 \\ 7 \\ 5 \\ 6 \end{bmatrix} \quad vt := \mathrm{Var}(t) = 1.593 \quad vc := \mathrm{Var}(c) = 2.5$$

$$k := \mathrm{length}(t) = 5 \quad sumvar := vt + vc \quad i := 1..k$$

$$p50 := \mathrm{qchisq}(.5, k) = 4.35146$$

$$ek := p50 \cdot sumvar = 17.810527$$

$$dev := \sum_i \left(t_i - c_i\right)^2 = 1.74$$

$$rp := \frac{(ek - dev)}{(ek + dev)} = 0.822$$

References for the Cattell Pattern Similarity coefficient r_p

Cattell, R.B. (1949). rp and other coefficients of pattern similarity. *Psychometrika*, 14, 279-298.

Cattell, R.B. (1969). The profile similarity coefficient, rp, in vocational guidance and diagnostic classification. *British Journal of Educational Psychology*, 39, , 131-142

Cattell, R.B. (1978). *The Scientific Use of Factor Analysis in the Behavioral and Life Sciences*. Plenum Press. ISBN: 0-306-30939-4

Cattell, R.B., Eber, H.W., and Tatsuoka, M.M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire*. IPAT.

Cattell, R.B., Coulter, M.A., and Tsujioka, B. (1966). *The taxonometric recognition of types and functional emergents.* In Cattell, R.B. (ed). The Handbook of Experimental Multivariate Psychology. Rand McNally.

A.1.6 Kernel Smoothed Distance

This coefficient is based upon a very simple idea that a distance function should be ***shaped*** in such a way that if the simple arithmetic unsigned difference between a person's attribute value and a target value was computed to be within a certain range, then the computed distance should reflect a very small distance, almost regardless of the actual distance. But, as that distance grows larger, then the computed distance should be accelerated in size. In short, an "**inertial**" effect was aimed for – translated into a distance metric. The coefficient itself is scaled as a measure of similarity, varying between 0 (maximal dissimilarity) to 100 (identity).

$$KSD = \frac{\sum_{i=1}^{n}\left[\frac{1}{s\sqrt{2\pi}}e^{-\left[\frac{(profile_{1i}-profile_{2i})^2}{2s^2}\right]}\right]\cdot\left(100\cdot\left(s\cdot\sqrt{2\pi}\right)\right)}{n}$$

where

$$s = \frac{range_i}{smoother}$$

$smoother$ = the "smoothing" parameter

$range_i$ = (maximum-minimum value) for case/variable i

$profile_{1i}$ = the attribute value i of n for profile 1

$profile_{2i}$ = the attribute value i of n for profile 2

n = the number of profiles

The key to using this coefficient is selecting an appropriate value of the smoother constant which produces the desired inertial effect. The selection choice is application-specific; the function in fact needs to be calibrated for each specific application, taking into account the costs and benefits of a sharper or smoother distance/discrepancy function. Person-target profiling applications are the most readily understood in this regard, where profile similarity can be adjusted to reflect only very close matches, with even "nearly similar" is reduced to "no-match" with even tiny departures of a person's profile from a target. Likewise, if a broad screen is required, then the smoothing function can be more gradual - providing a kind of "plateau" effect around the target value, before the discrepancy between person and target is accelerated by the non-linear function.

In essence, this coefficient needs to be "calibrated", to match the "by eye" judgment of the user. That is, when plotting two profiles, or when deciding whether two values are to be adjudged similar to one another given the range of the rating scale being used, it is the user who has to decide when two values are to be adjusted "similar", and not the "mathematics" . The KSD coefficient is sensitive only to

magnitude discrepancy (not monotonicity). It also takes into account the range of ratings or score values from which the person-target, or rater reliability ratings are drawn.

For example, consider the comparison of two sets of scores from two Raters ...

Rater 1	Rater 2
3	4
4	3
3	4
4	3
3	4
4	3
3	4
4	3
4	4
4	4

Such data will produce highly negative Pearson and Intraclass correlations.

Assuming a rating scale range of **1-5**, the KSD coefficient with smoother factor of 5 is: **0.57**. If we assume the scores are drawn from a range **1-20**, then the KSD coefficient is **0.97**. Clearly, the design of the coefficient introduces the element of relativity; a score discrepancy of 1 looks reasonably important when the range is just 1 to 5, but trivial when the range is 1-20.

If we apply the coefficient formula to the data in the table, with a KSD smoother value of 5,

Rater 1	Rater 2	KSD
3	4	45.783
4	3	45.783
3	4	45.783
4	3	45.783
3	4	45.783
4	3	45.783
3	4	45.783
4	3	45.783
4	4	100
4	4	100

the average of these 10 rater pairs is 56.63% similarity, or using a 0-1 coefficient scaling, 0.57.

Incidentally, the Pearson correlation for these data is – 0.67, ICC model 2 = – 0.80, and model 3 = – 0.67, while the Gower is +0.80, and Double Scaled Euclidean-Similarity = +0.78.