Person-Target Profiling for Selection & Recruitment

Part 1: Introduction

Paul Barrett

Director of Research and Development

Psytech International (UK) Ltd.

paul.barrett@psytech.co.nz

Adjunct Professor of Psychometrics & Performance Measurement
University of Auckland
Faculty of Business

paul.barrett@auckland.ac.nz

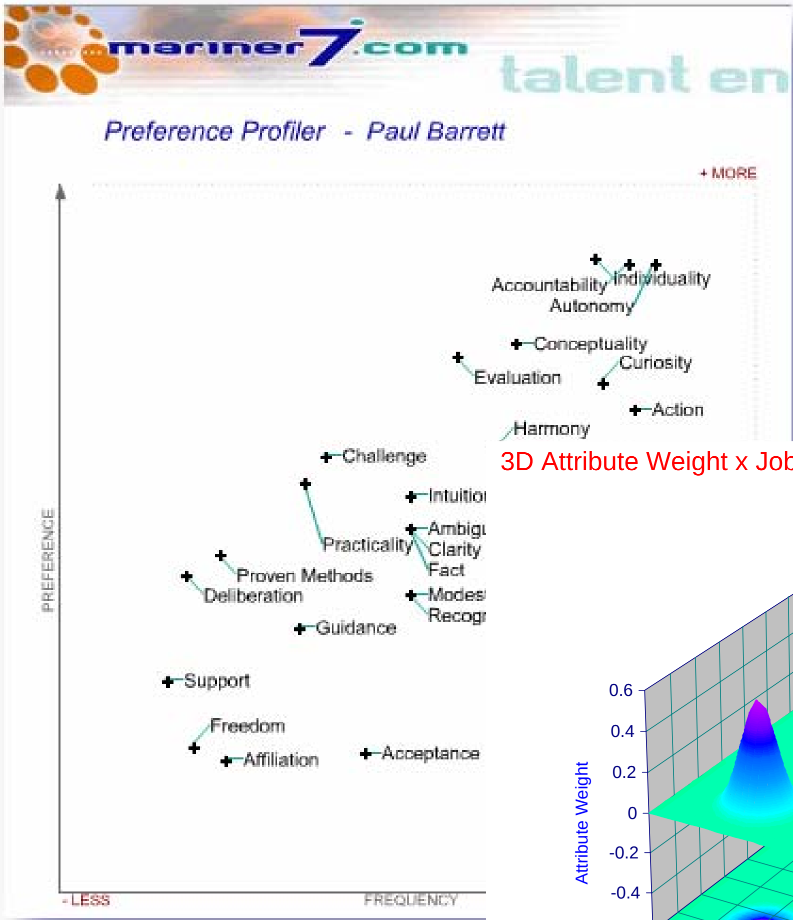
Adjunct Associate Professor of Psychology

University of Canterbury
Department of Psychology

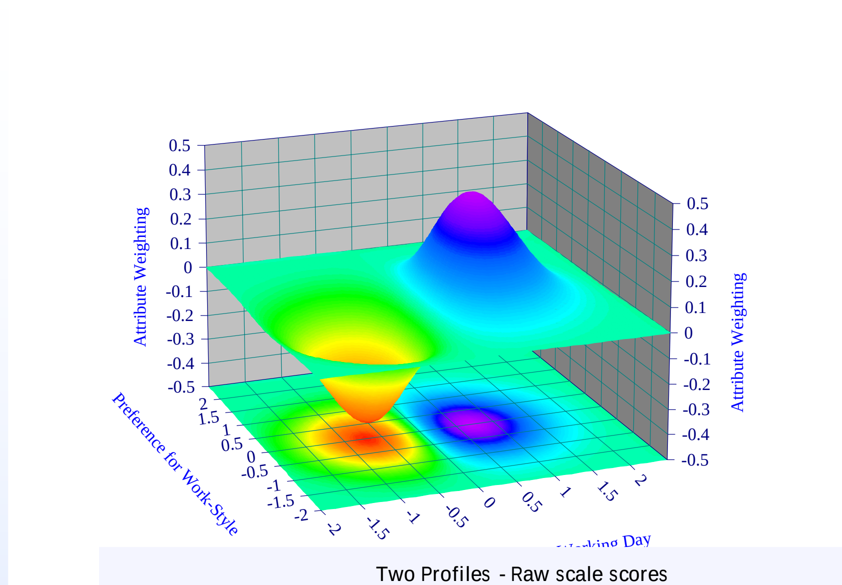
paul.barrett@canterbury.ac.nz

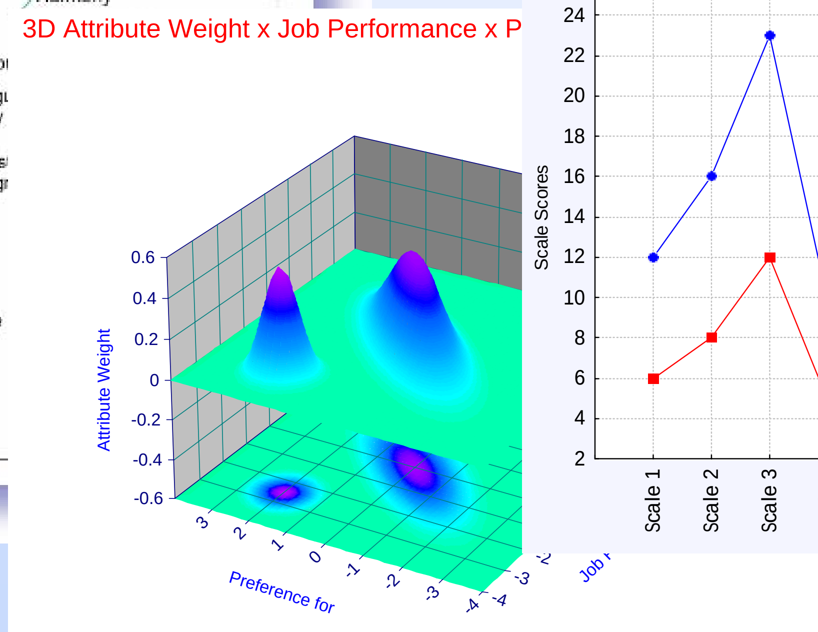
Web: www.pbarrett.net

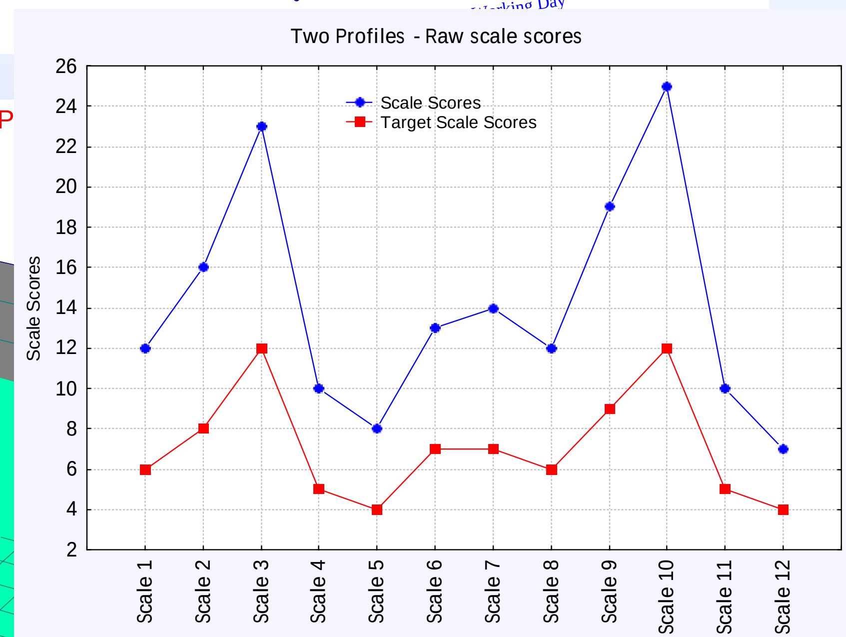
The Basics



3D Profile: showing bipolar nonlinear categorical scoring







The Profile: A Definition

A profile is defined by Collins English Dictionary (1991,3rd edition) as: "a graph, table, or list of scores representing the extent to which a person, field, or object exhibits various tested characteristics or tendencies". This general definition characterises the typical usage of the term in I/O psychology, where scores on tests and/or items of information are used to characterise persons, jobs, or groups of either.

More specifically, a profile may be defined as an outline or shape formed by plotting magnitudes for an individual, group, or job, into a 1, 2, or 3 dimensional space.

Some Terminology

The constituent attribute values that define a profile may be viewed as a "**target**" profile, against which other profiles will be compared, or as a "**comparison**" profile which is the profile which is to be compared to the target.

Some Terminology

Cronbach and Gleser (1953) introduced three terms to describe a profile:

Elevation: the mean of all scores within a single profile.

Scatter / Variability : the square root of the sum of squares of a single profile's deviation scores about the Elevation for that profile. Essentially the standard deviation of scores within a profile, multiplied by the square root of the number of attributes constituting the profile.

Shape : the residual information left in each score of a profile, after equating for the elevation and scatter indexes by subtracting out the elevation and dividing the resultant deviation score by the scatter value.

Some Terminology

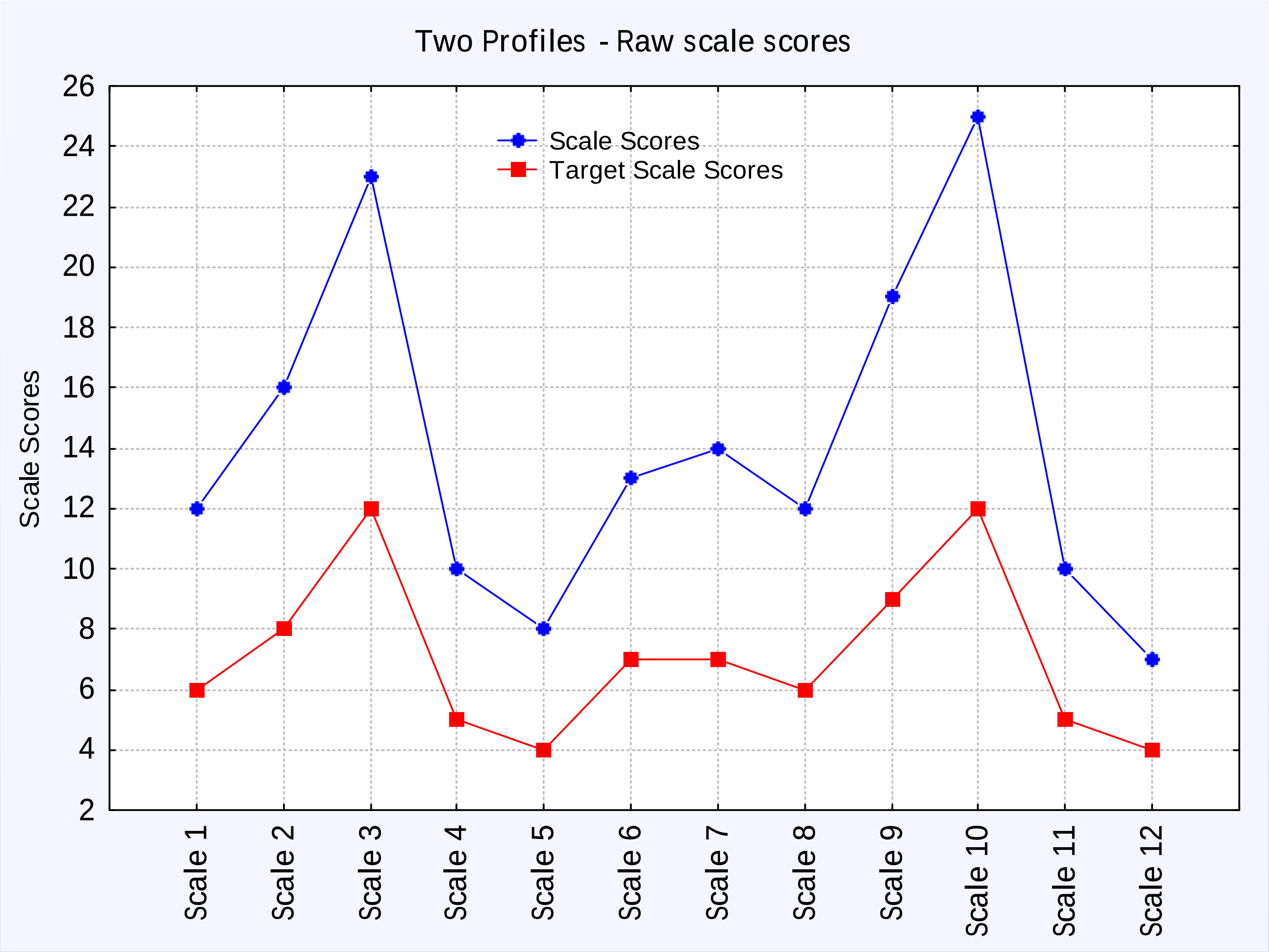
Given a profile (1..*n*) of scores (*x*)...

$$Elevation = \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

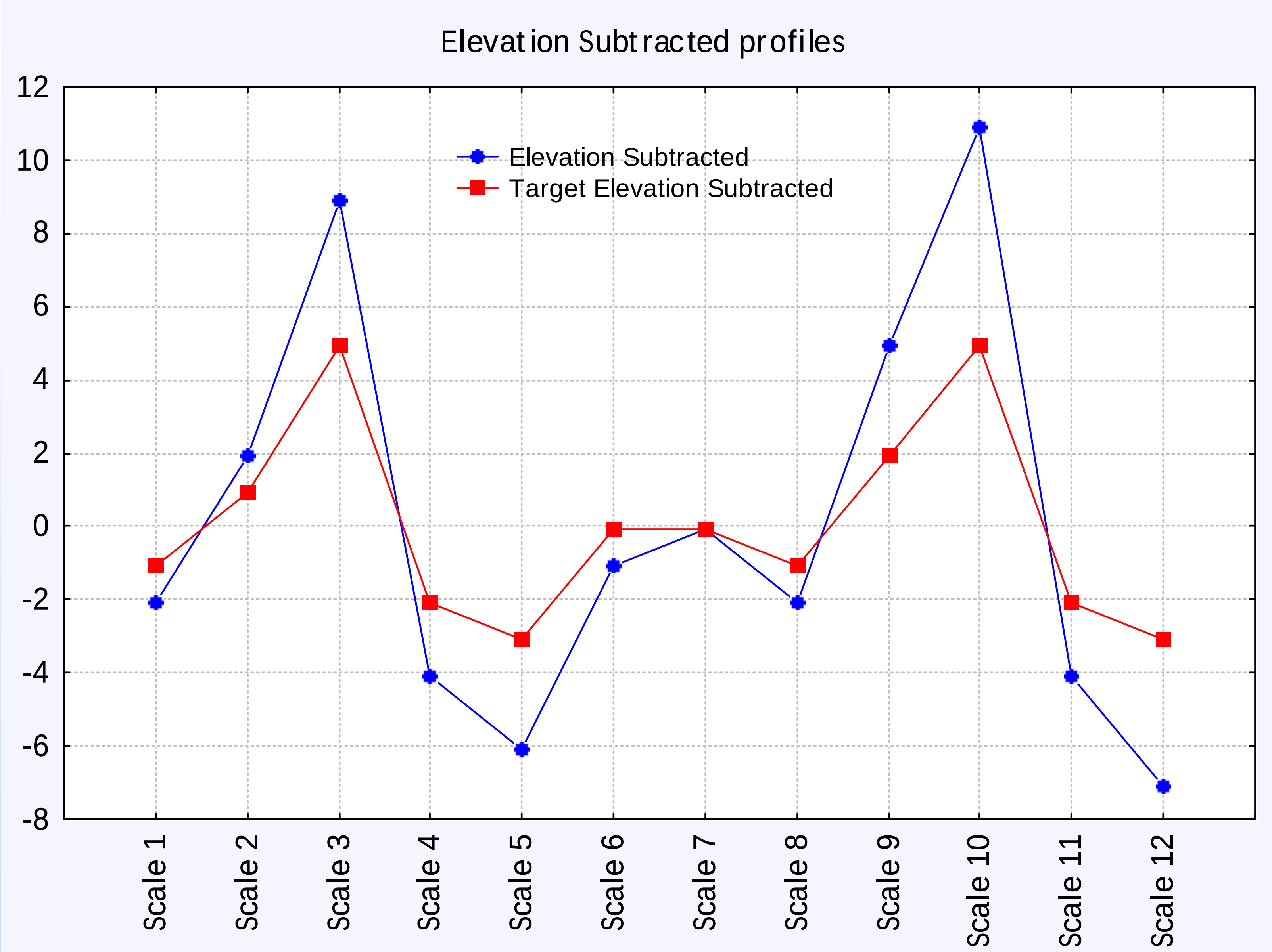
$$Scatter = \left(\sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}} \right) \cdot \sqrt{n}$$

$$Shape = \frac{\left(x_i - \overline{x}\right)}{Scatter}$$

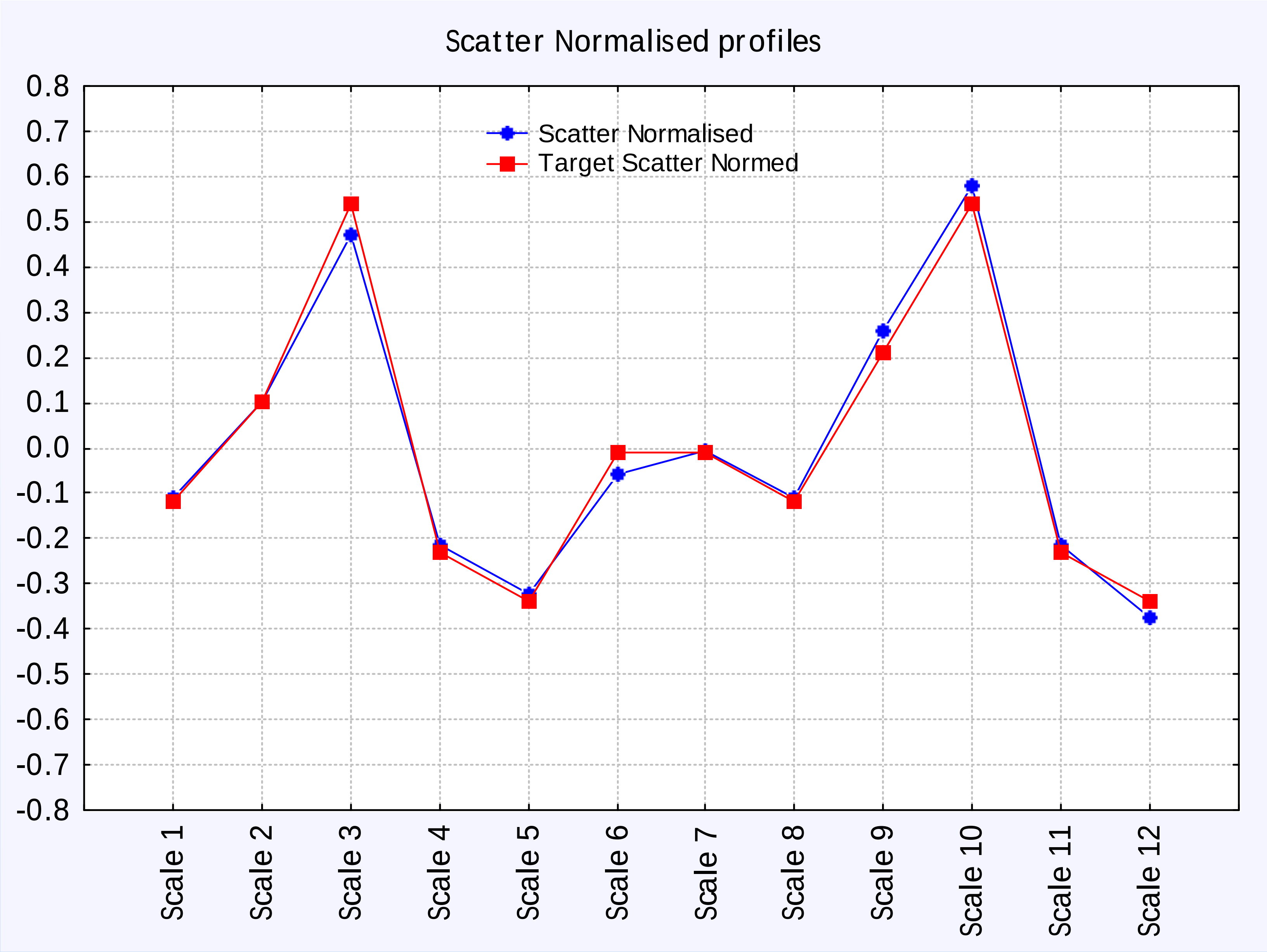
A simple example: two raw-score profiles



Now with their respective Elevation parameter subtracted from each



Now we divide through the de-elevated profiles by their Scatter



Sensitivity of certain similarity coefficients to the transformations

And these are some profile similarity coefficients computed from the three previous plots …

	Pearson r	ICC Model 2	ICC Model 3
Raw profile comparison	0.99	0.35	0.78
De-Elevation comparison	0.99	0.79	0.78
De-Scatter comparison	0.99	0.99	0.99

Why Profile?

- For summary descriptive purposes, making largely qualitative statements about individuals, jobs, etc. *For example, company X tries to ensure that all its managers demonstrate significant customer focus.*
- For computational purposes, where the profile will be used as a "target" against which a quantifiable "distance" or "similarity" index will be computed. *For example, company X ensures that all its managers demonstrate a minimum measured level of customer focus.*
- We are concerned with the latter.

Why Profile?

✪ For reasons of efficiency, predictive accuracy, and objectivity. Let us not forget the evidence of Grove and Meehl (1996), Grove et al (2000), and Swets et al (2000)

Grove, W.M., & Meehl, P. (1996) Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures. *Psychology, Public Policy, and Law*, 2, 2, 293-323.

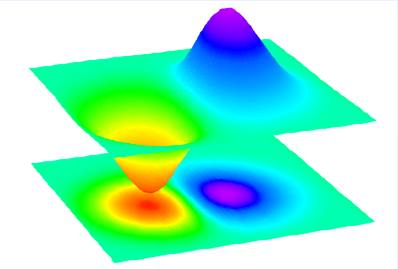
Grove, W.G., Zald, D.H., Boyd, S.L., Snitz, B.E., & Nelson, C. (2000) Clinical vs mechanical prediction: a meta-analysis. *Psychological Assessment*, 12, 1, 19-30.

Swets, J.A., Dawes, R.M., & Monahan, J. (2000) Psychological Science Can Improve Diagnostic Decisions. *Psychological Science in the Public Interest*, 1, 1, 1-26.

Current examples of Commercial Applications of Profiling

The Psytech International (and Psytech SA) GeneSys™ 1-D Candidate Profiler System

The Mariner7.com, Staffcv Inc., 2-D Graphical Profile



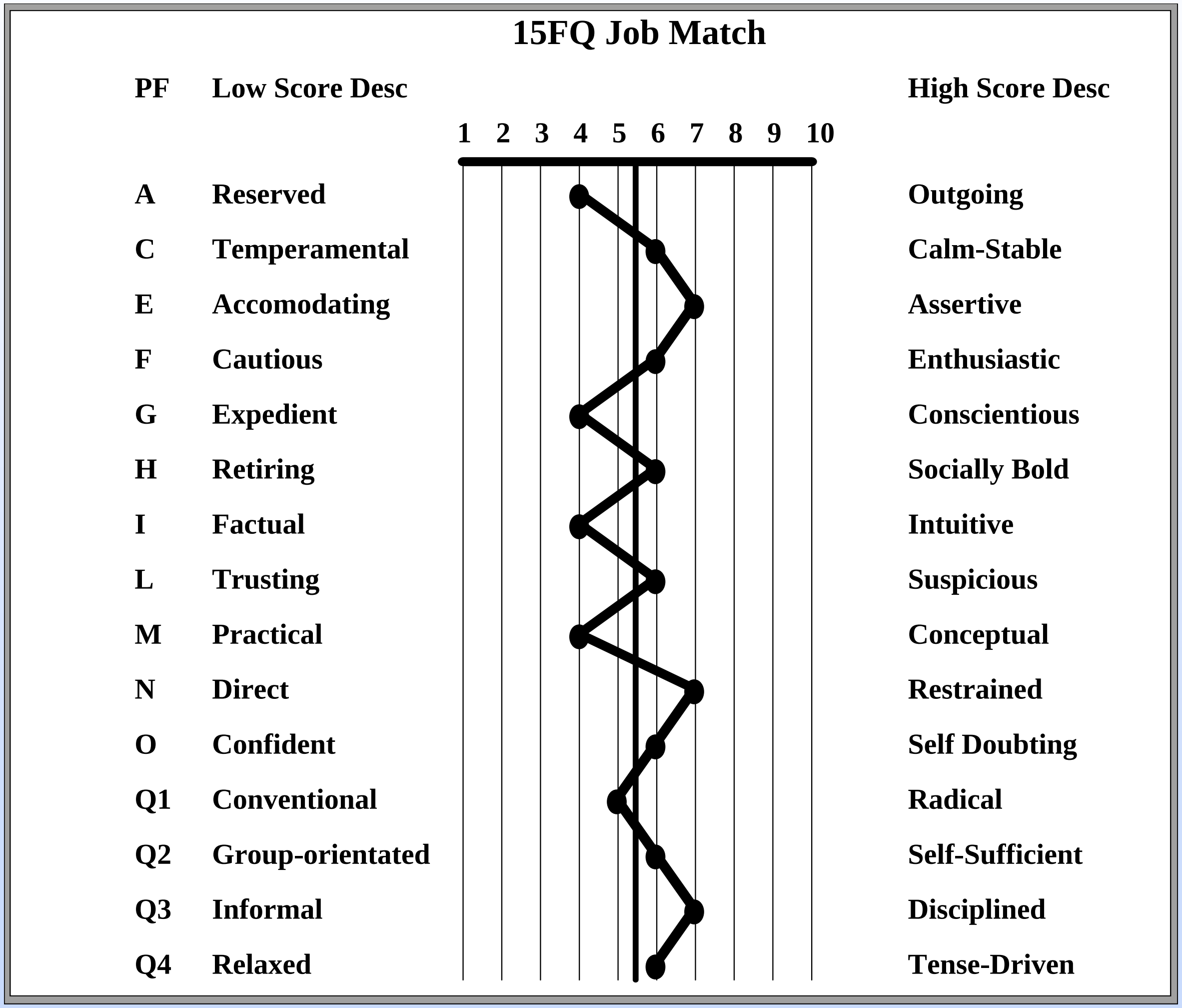
The 3D profile – independent work "in progress" by me!

The GeneSys™ Candidate Profiler System

A one-dimensional computer-based profiling system that is an exemplar of all such systems currently in use. Basically, the system uses the data acquired by the administration of its bank of tests (ability, personality, interests, motivation, work-styles etc.) to allow the user to develop profiles for categories of individuals and/or set up "ideal profiles". Then, all existing records may be compared to this profile, or a new individual can be scored for similarity. The system currently uses Cattell's Pattern Similarity Coefficient as its index of similarity.

The GeneSys™ Candidate Profiler System

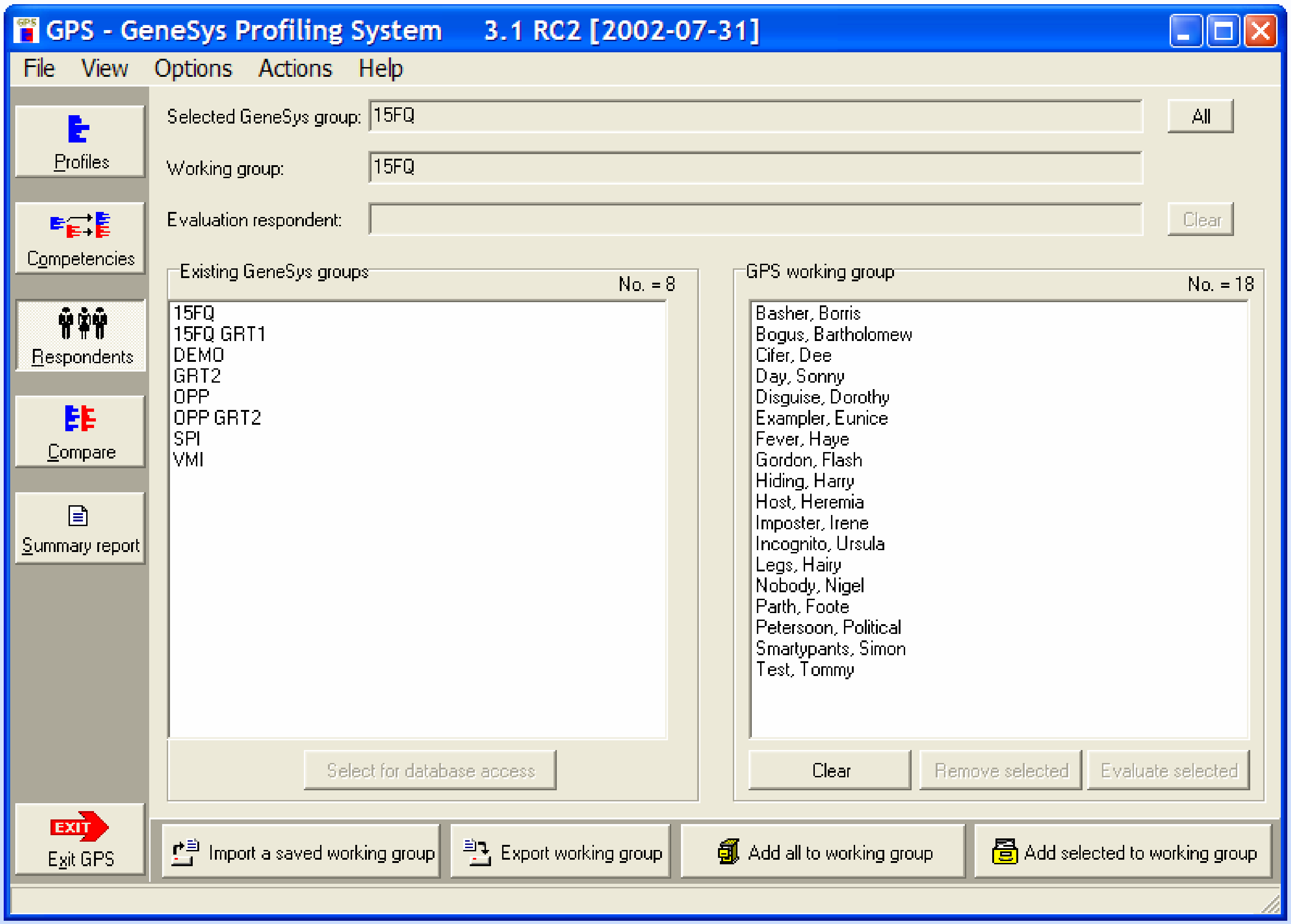
A typical "ideal profile"



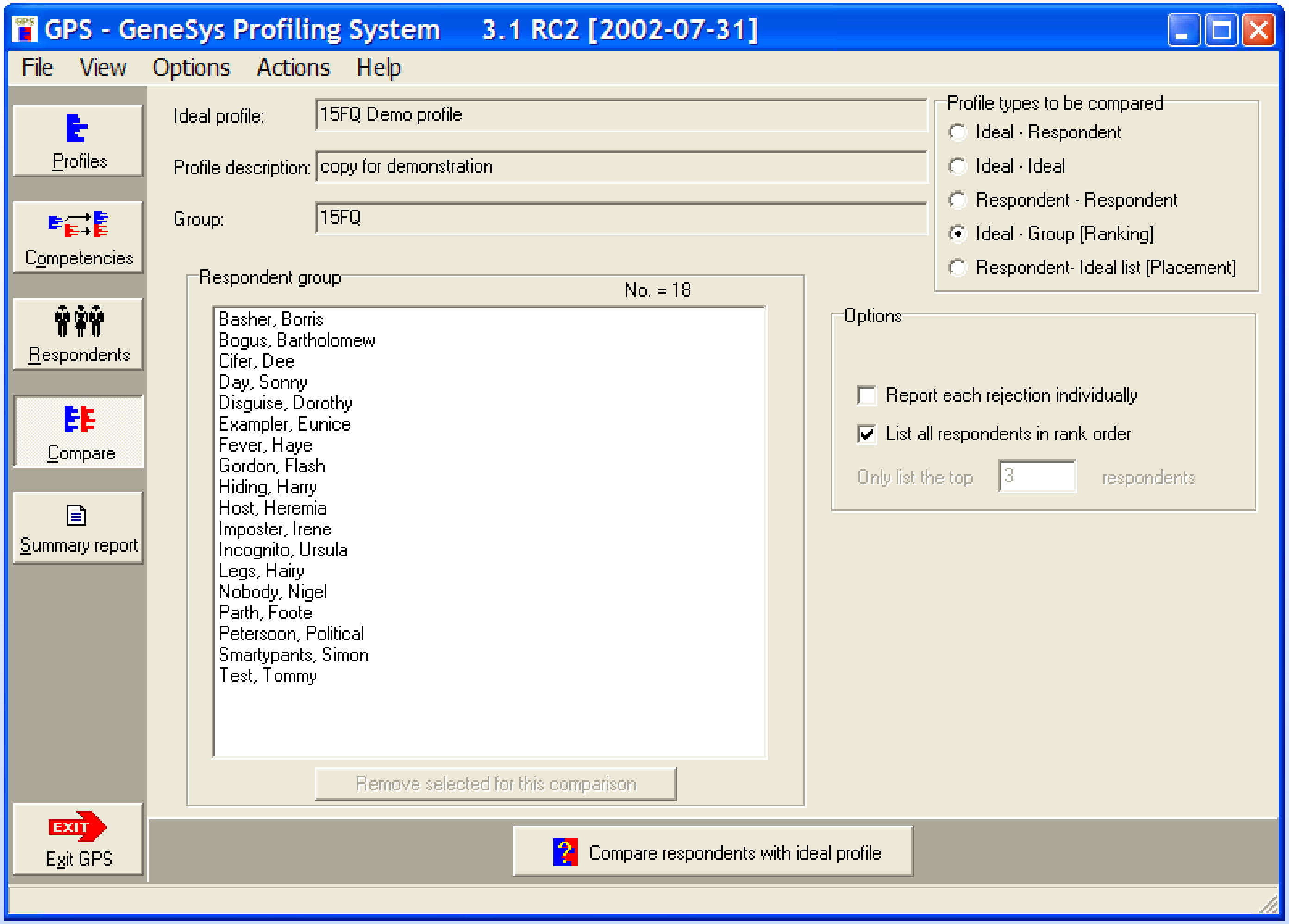
The GeneSys™ Candidate Profiler System

Having set this profile up – Genesys™ will then search through a group of chosen individuals and rank order them in terms of similarity to the ideal target profile. You can then "drill-down" to look at a profile chart of a best-fitting individual's profile comparison to the target profile.

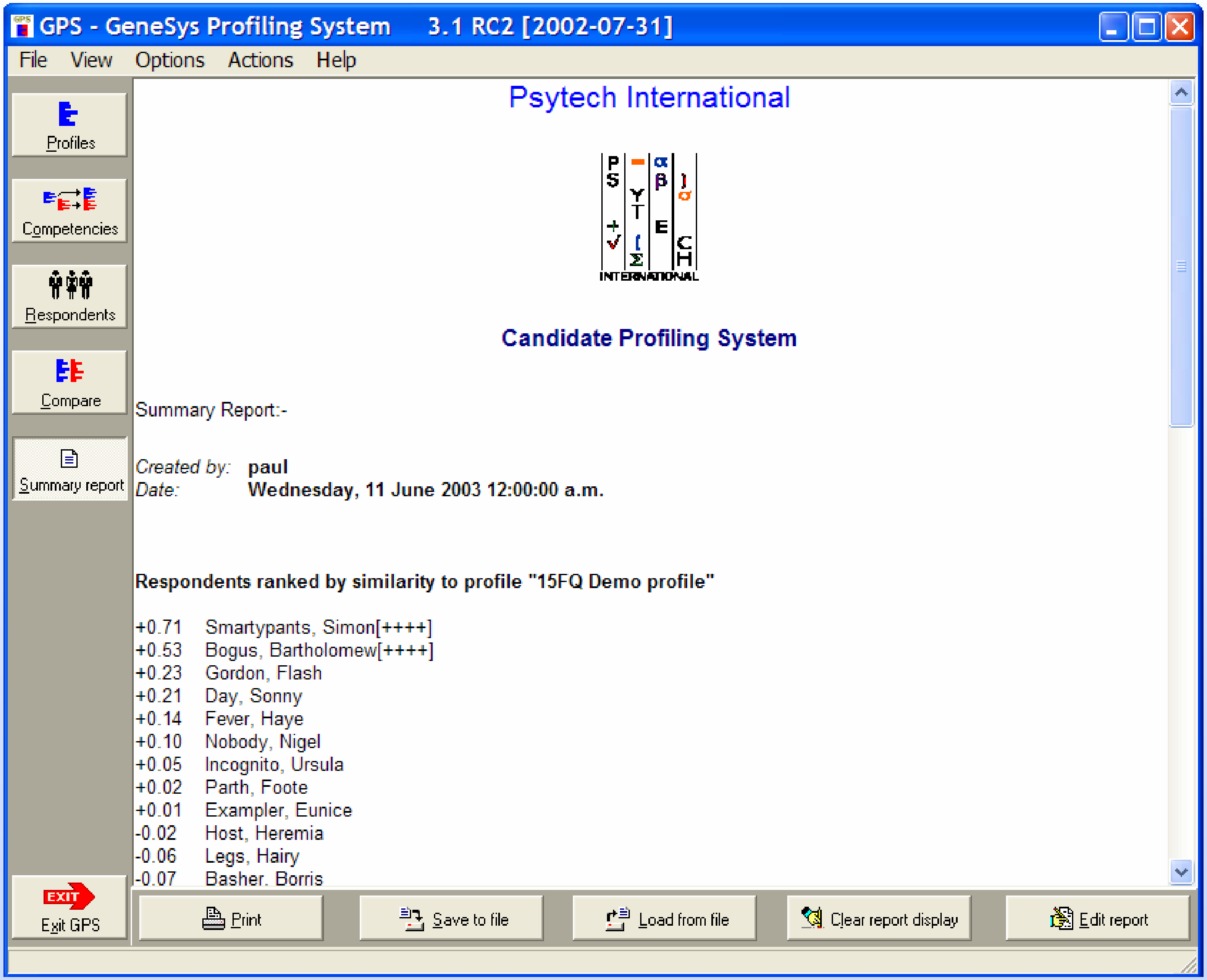
The GeneSys™ Candidate Profiler System



The GeneSys™ Candidate Profiler System

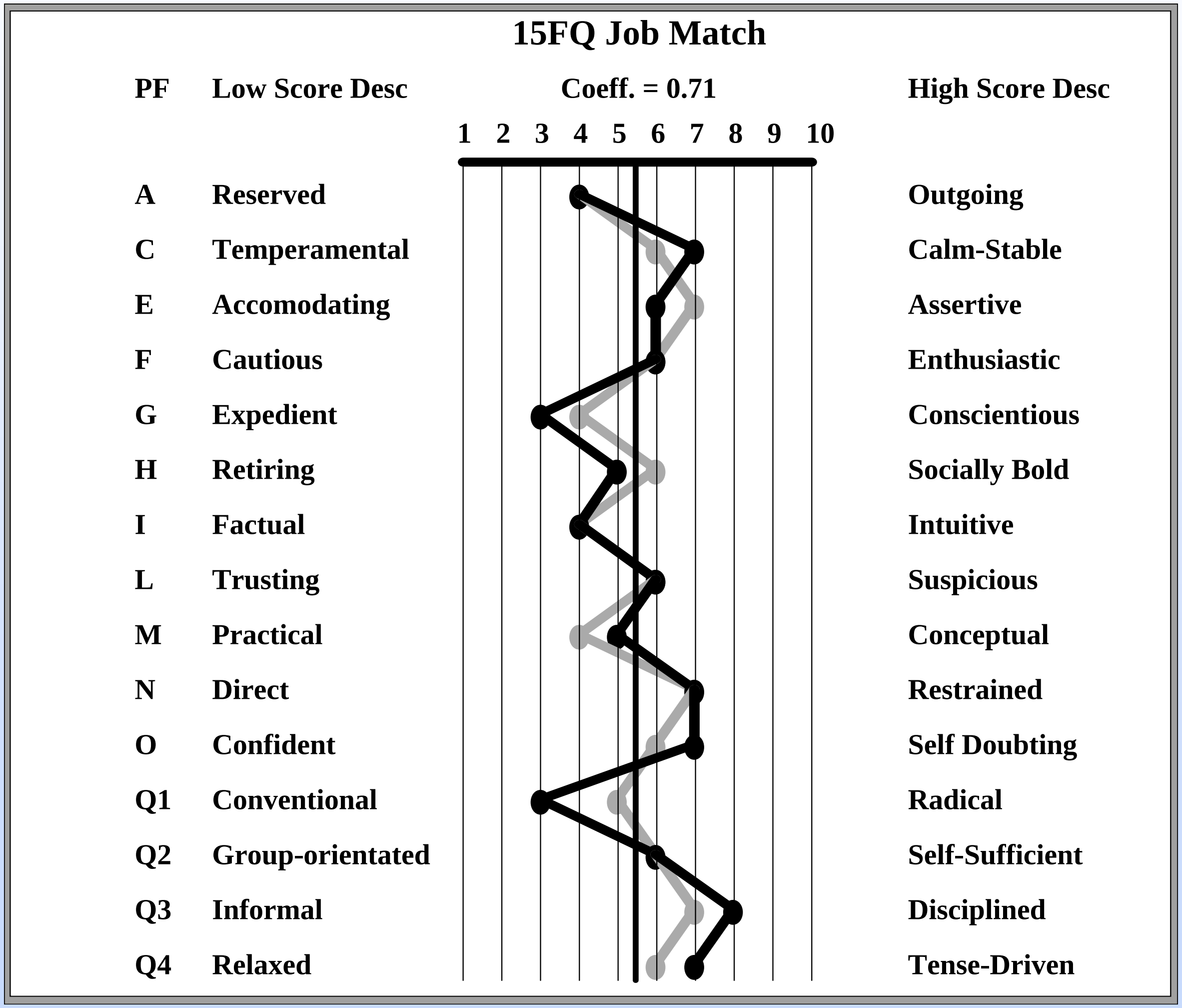


The GeneSys™ Candidate Profiler System



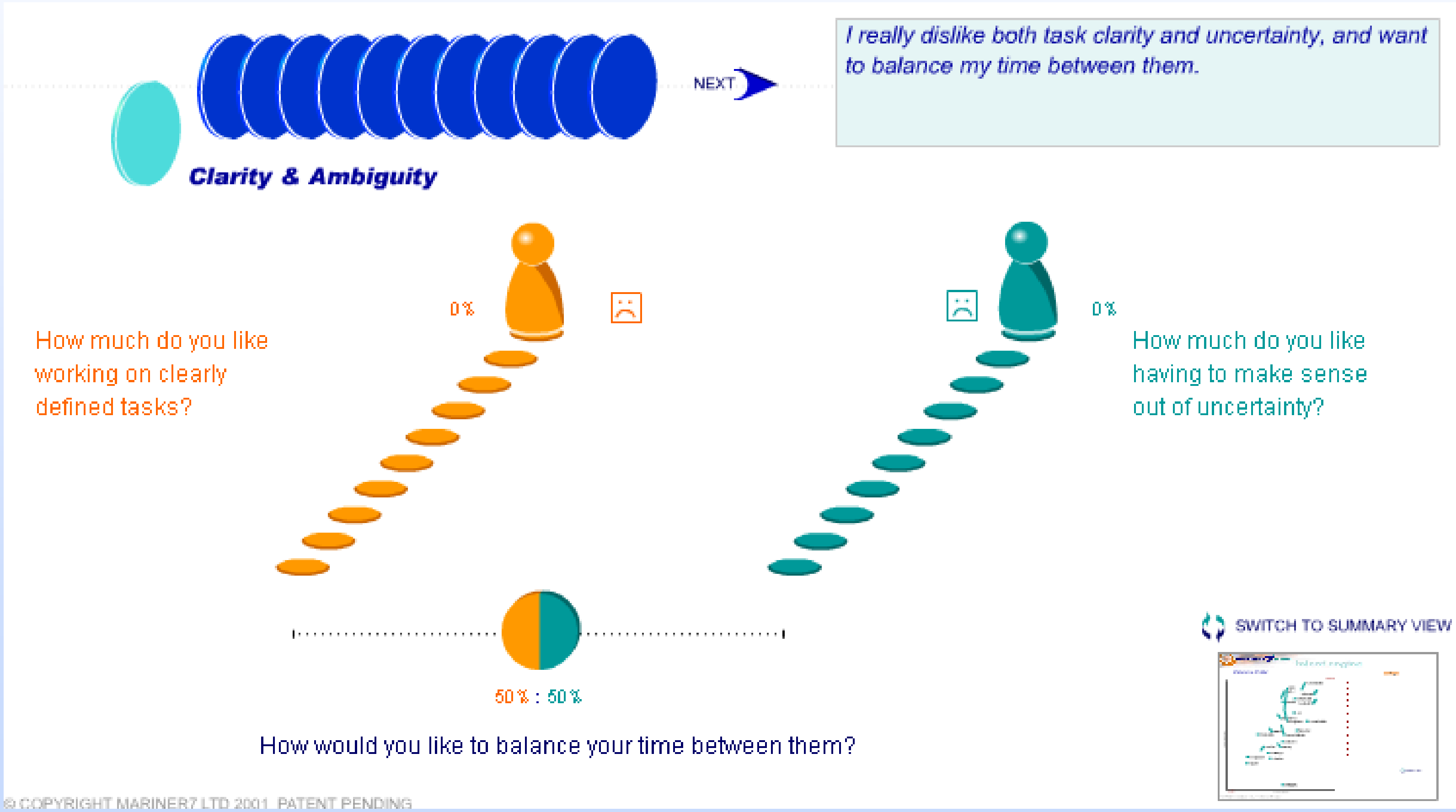
The GeneSys™ Candidate Profiler System

Simon Smartypants' profile comparison

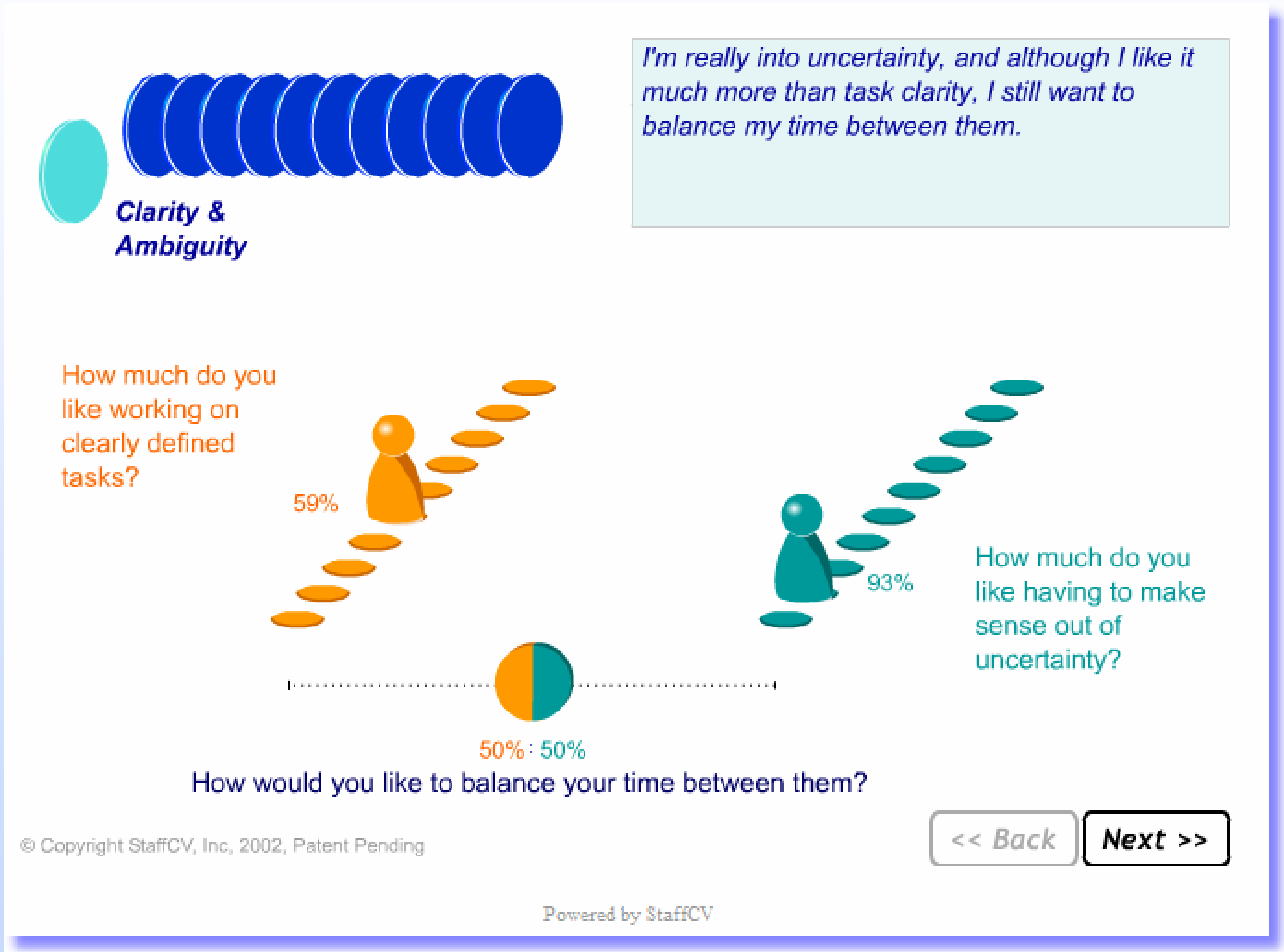


The Mariner7-Staffcv Graphical Profiler

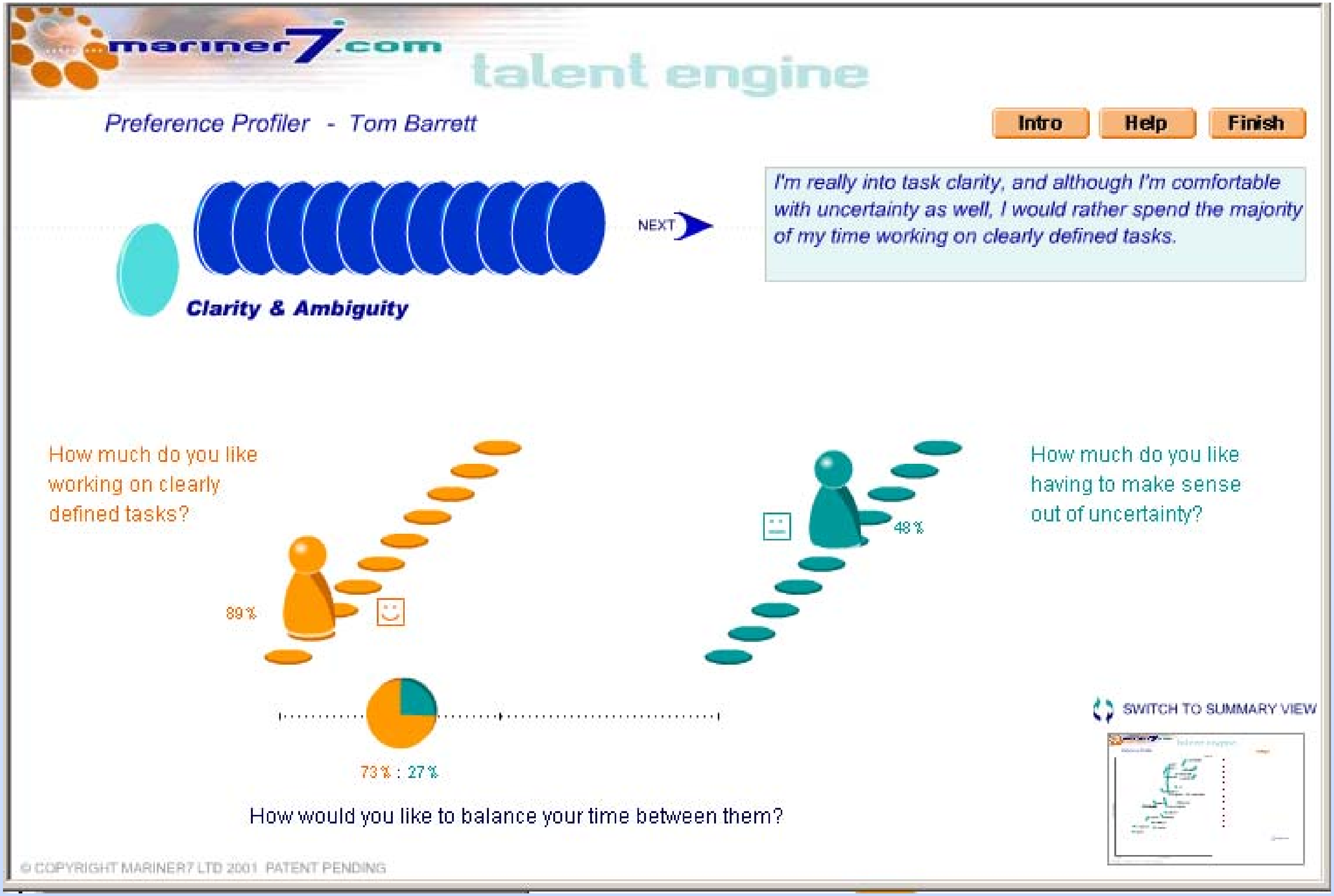
The user is required to indicate their work-preferences and the frequency with which they like to engage in them during a working day.



The Staffcv profiler format …

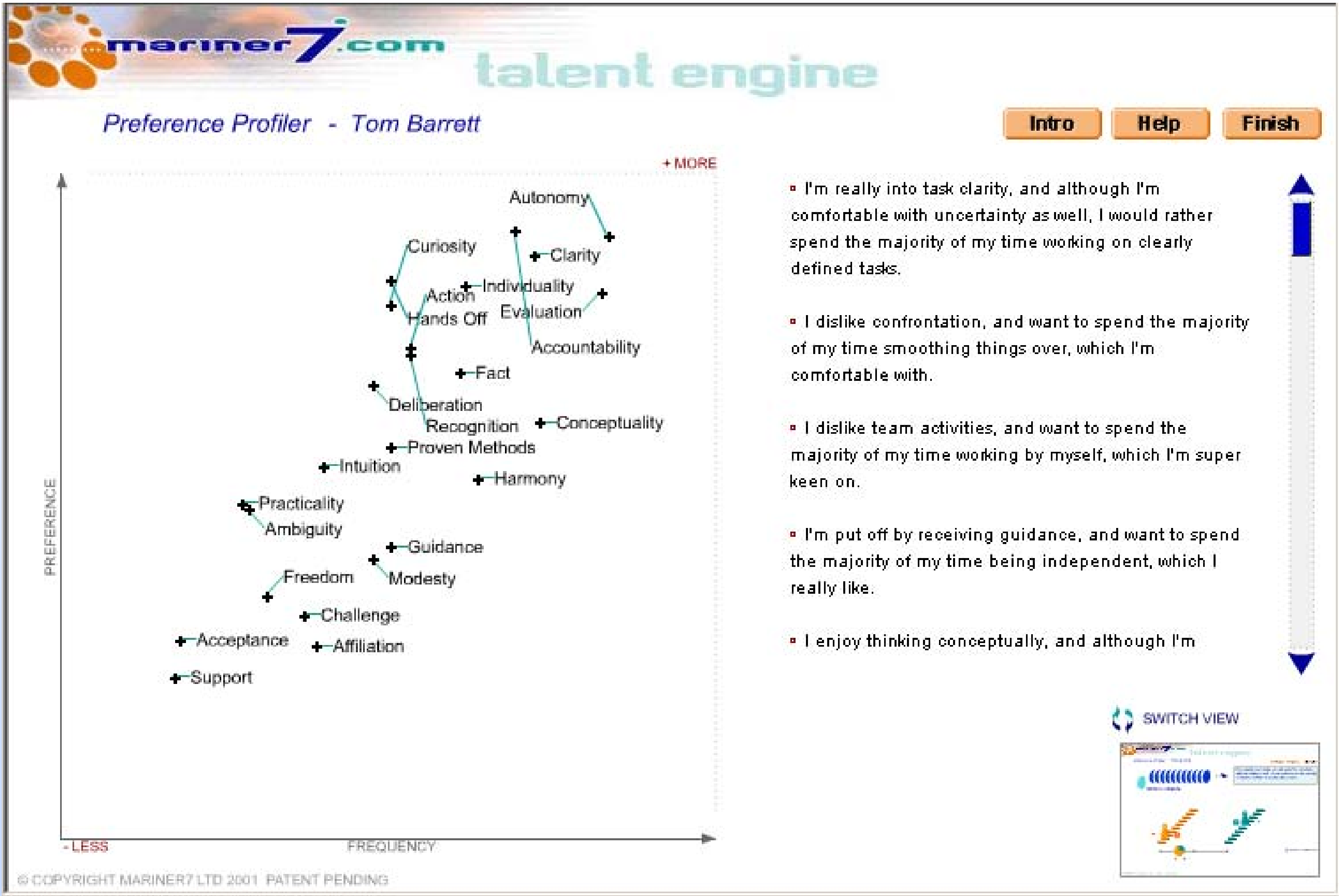


The Mariner7-Staffcv Graphical Profiler

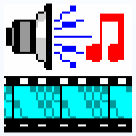


The Mariner7-Staffcv Graphical Profiler

The 2-Dimensional Profile



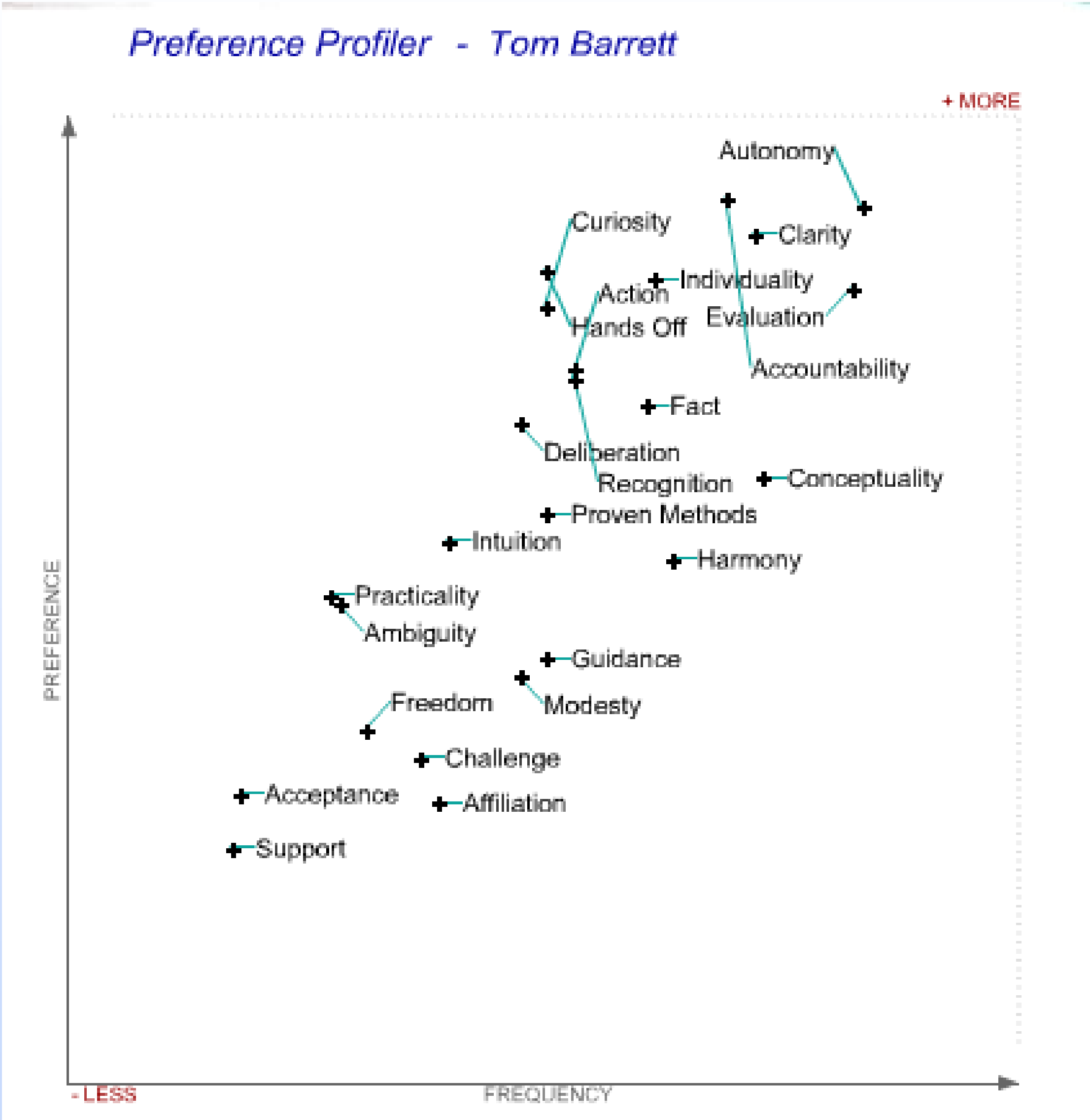
The Mariner7-Staffcv Graphical Profiler

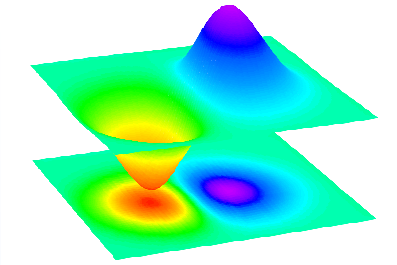


The movie!

The Mariner7-Staffcv Graphical Profiler

This is the 2-D profile actually used for comparison purposes



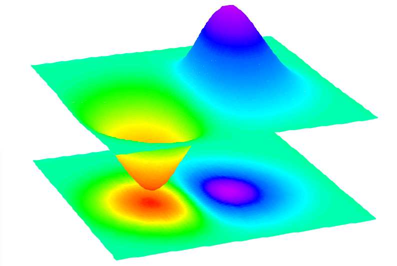


The 3-Dimensional Profile

Work in progress – me!

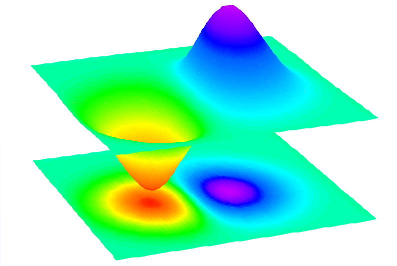
Simple linear methods of profile comparison, as in the conventional 1-D profiling techniques, all rely upon some kind of linear differencing/covariation estimate between target and comparison attribute, with perhaps some "non-linearity" introduced as part of the "weights" applied to each profile attribute in a composite profile (as in the Genesys Profiling System).

But …



The 3-Dimensional Profile

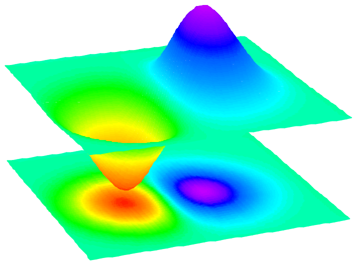
But, what if we choose to penalise or reward a distance from target to comparison attribute score using a non-linear function, that is smooth over the ranges we consider "relevant", and allows for accelerated incremental distance and weights that vary between positive and negative attribute effects?



The 3-Dimensional Profile

Further, as in the case of the Mariner7/Staffcv 2-D profile, what if we also wish to differentially weight the attributes prior to their entry into the overall 24 attribute profile? To do this, we need to work in 3 dimensions …

An example …

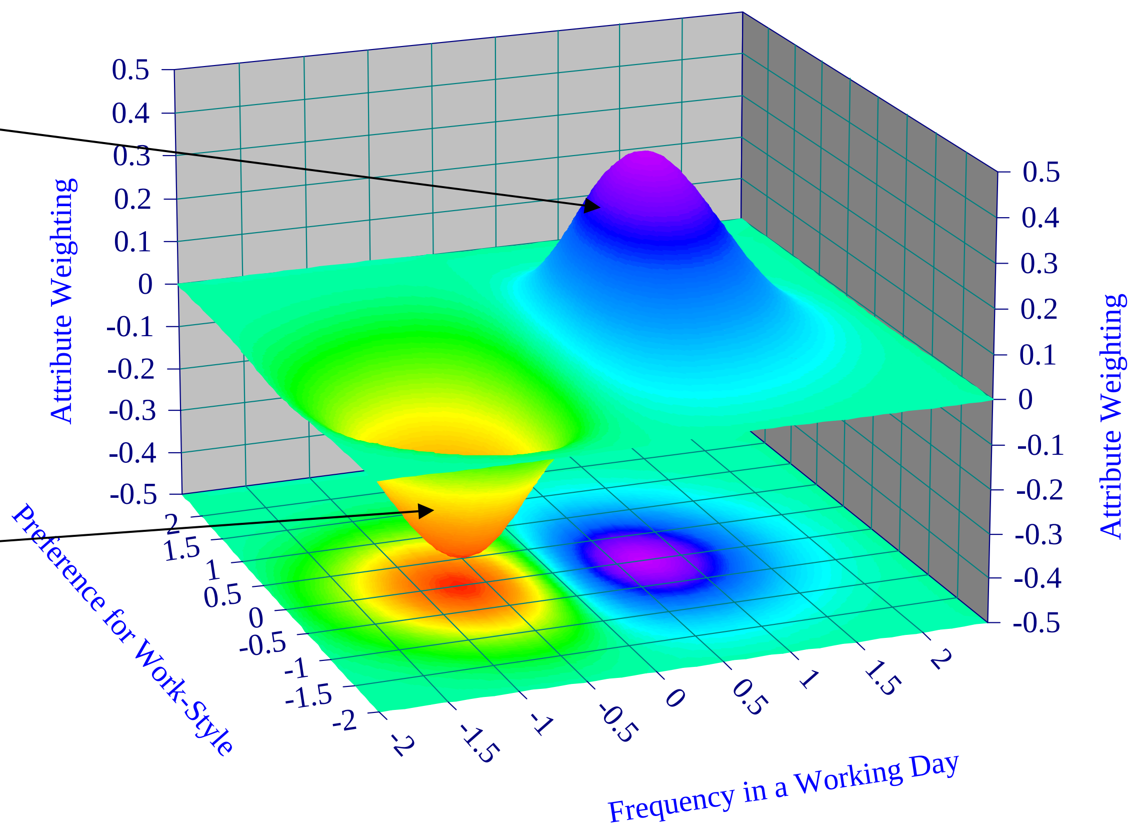


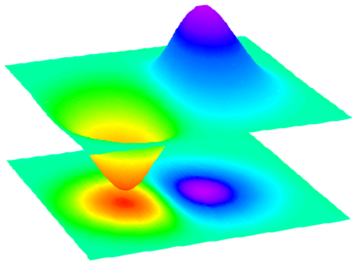
The 3-Dimensional Profile

3D Profile: showing bipolar nonlinear categorical scoring

person A
score - positive
weight into a
composite
profile

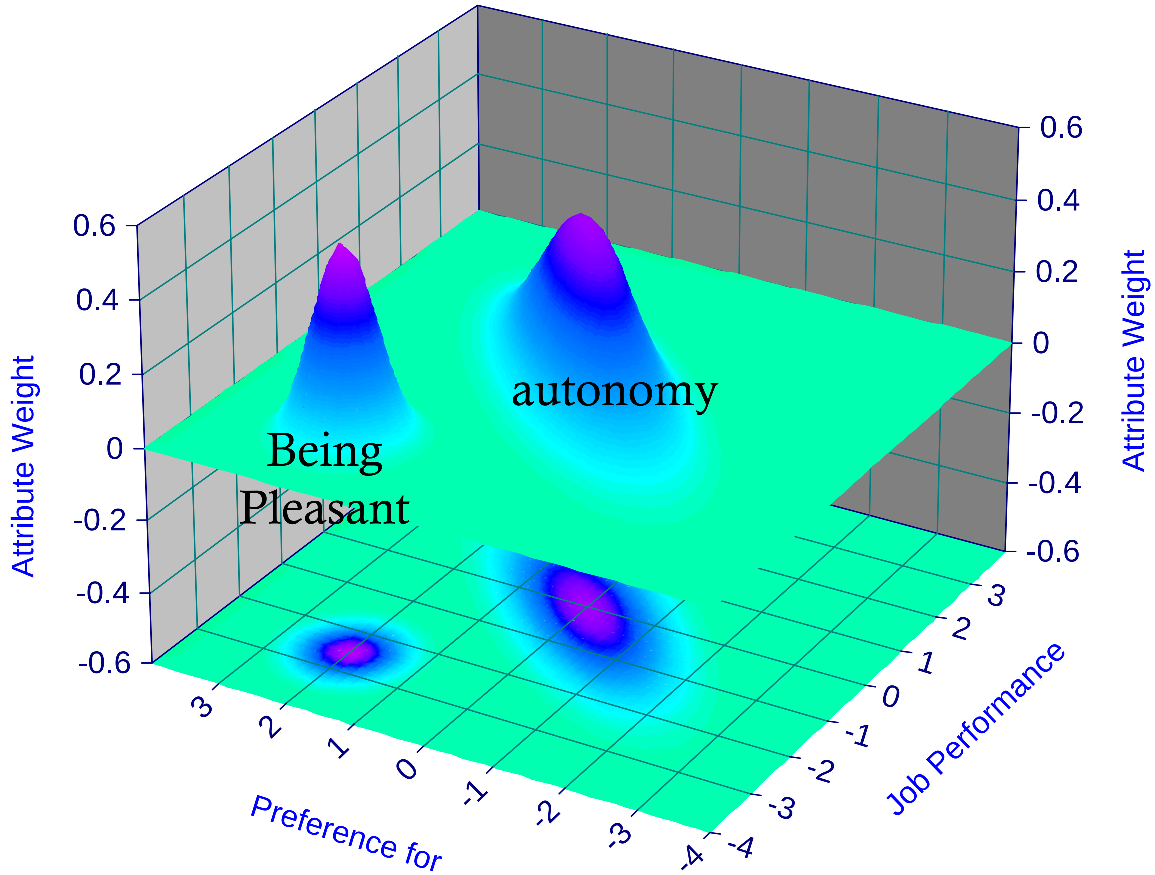
person B
score - negative
weight into a
composite
profile





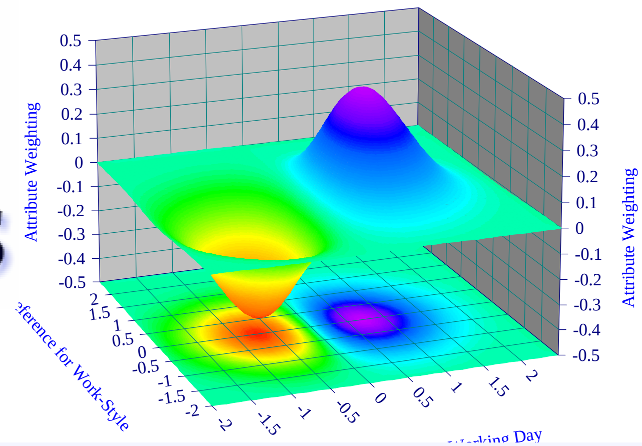
The 3-Dimensional Profile

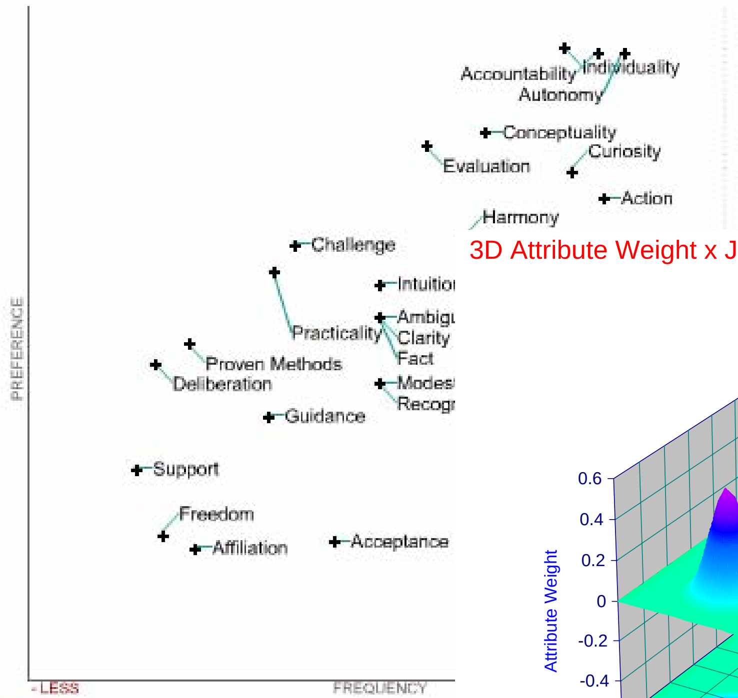
3D Attribute Weight x Job Performance x Preference for Work-Type



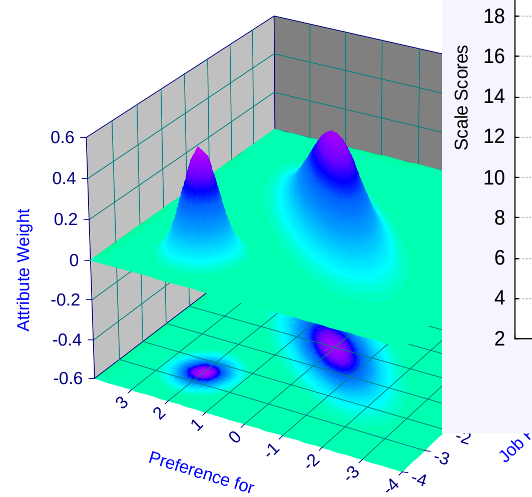
The gap between expectations and delivered solutions

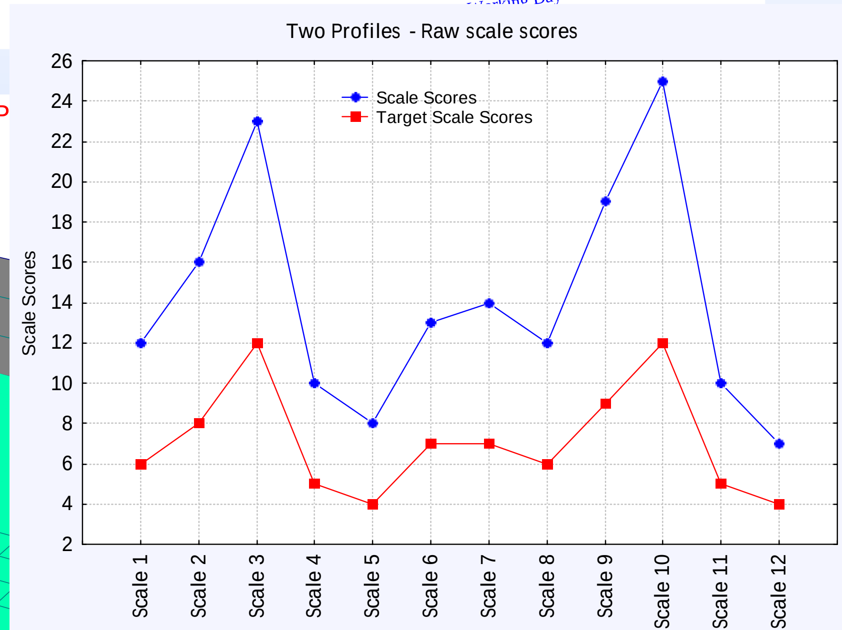
3D Profile: showing bipolar nonlinear categorical scoring





3D Attribute Weight x Job Performance x P





Reasons for the gap .1

❶ The standard linear methods for coordinate distance and pattern similarity calculation are not particularly suited to profiling (e.g. euclidean distance, pearson correlation, congruence calculations). Further, no "sensitivity analysis" is ever undertaken to determine the sensitivity of coefficient size in relation to aspects of profile comparison disparity (e.g. the change in size of a summary matching coefficient relative to the actual discrepancies in the two profile vectors)

Reasons for the gap .2

❷ The very restricted nature of the data that enters a profile. i.e. Generally, this is restricted to small subsets of all available data for an individual, such as biodata or performance data or 360 degree data or psychometric test data. It would be fair to say that the majority of commercial HR profiling is done solely with psychometric test scores. No system yet copes with all possible data sources – apart from one constructed empirically via data mining and using some unique blend of sophisticated classification techniques.

Reasons for the gap .3

❸ HR professionals, recruitment consultants, and many psychologists are generally dubious of “statistical prediction rules” or profile matching. This is partly due to:

✪ the apparent poverty of domain breadth of many kinds of target profile.
✪ the insensitivity (or sometimes unknown sensitivity) of the summary coefficient to what may be significant discrepancies between certain attributes in a multi-component profile with many attributes.
✪ the lack of empirically validated or computationally modelled data to support the use of a target profile.

Reasons for the gap .4

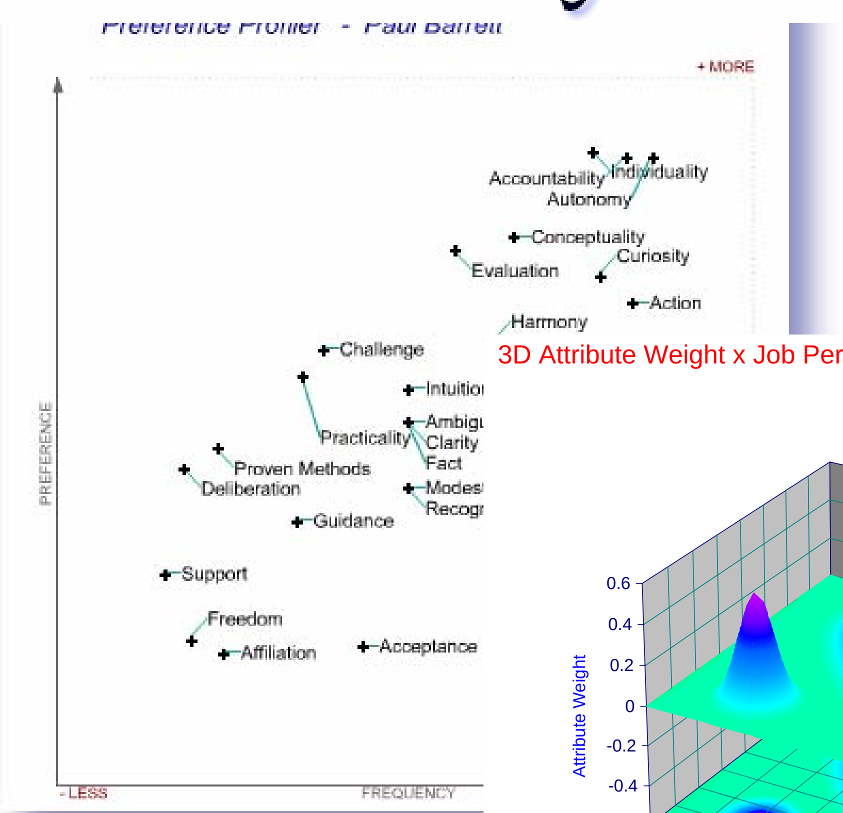
❹ There is a real deep-seated philosophical concern as to whether any target profile can ever be constructed that would consistently and reliably determine "successful" vs "not so successful" employees. This is especially the case when using a single "star performer" profile as a target profile, or even a group of employees who all do the same job etc. – and constructing the "optimal" profile from them so as to hire "more like them" in the future.

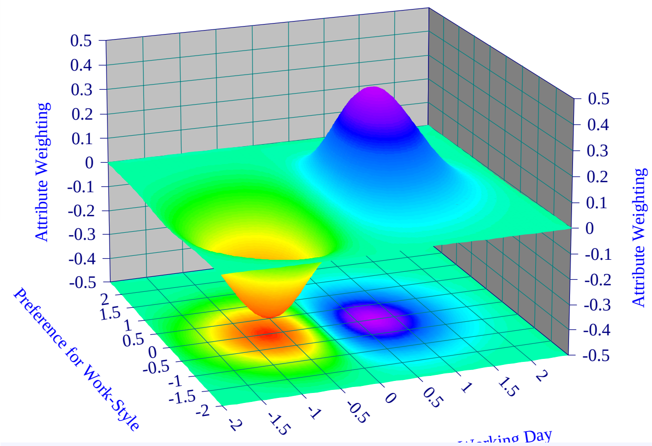
Reasons for the gap .5

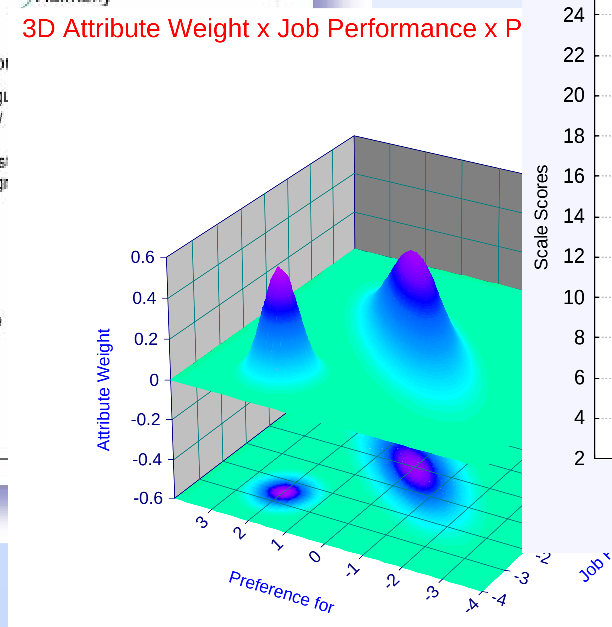
❺ There is a very real difficulty concerned with exactly how to define an optimal target profile in the first place, let alone compare others against it! This is where many fledgling profiling methods falter initially. In some respects, this is probably the most significant and invariably fatal "hiccup" on the path between "sales-person induced customer expectations" and "reality".

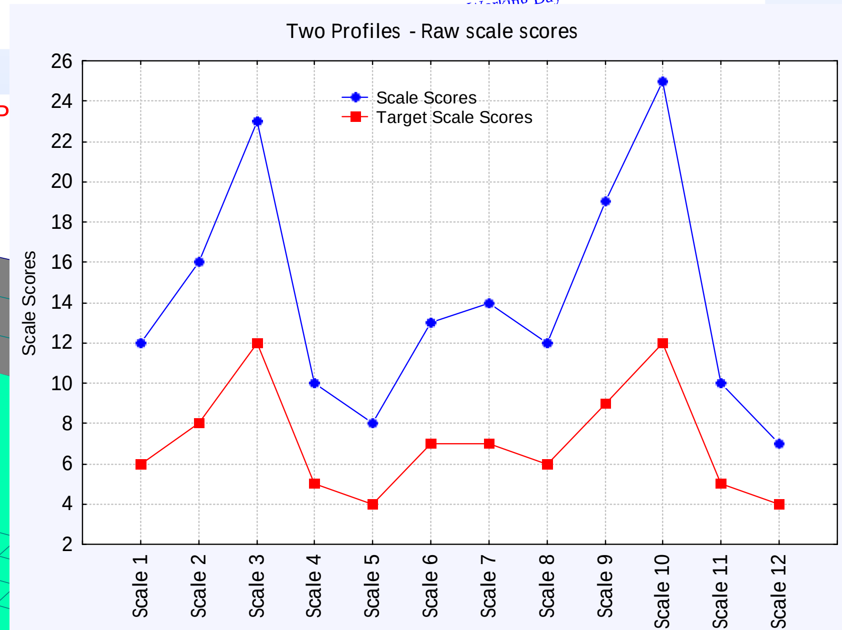
3D Profile: showing bipolar nonlinear categorical scoring

The Target Profile in its many forms









The Target Profile

- ✪ Hypthesised, "**ideal**" profile
- ✪ Single empirical profile – the "**star performer**"
- ✪ The multiple **team profile** - "single" profiles for each team member within a "perfect team"
- ✪ Homogenous empirical profile – the construction of a single "**homogenous group profile**" from a group of individuals who are all adjudged to be homogenous with respect to some external attribute (such as their job, level within a company, gender, team position etc.)

The Target Profile

Further, the target profile may consist of:

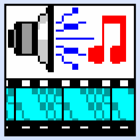
- Equal-interval scores
- Ranked, ordered attribute values
- Categorical classes
- Correlated constituent components
- Independent constituent components
- A mixture of all of the above!

Which will very meaningful when it comes to selecting an index which will be proposed as indicating distance or similarity.

The Ideal Profile

Quite simply, the profile that is created subjectively.

- ✪ Use a package like Visio to “play” with a 1-D profile in order to setup the desired levels.
- ✪ Use a 1-D profiler like the **Barrett Personality Profiler** to construct a desired profile *(installation demo program available on the Workshop CD)*

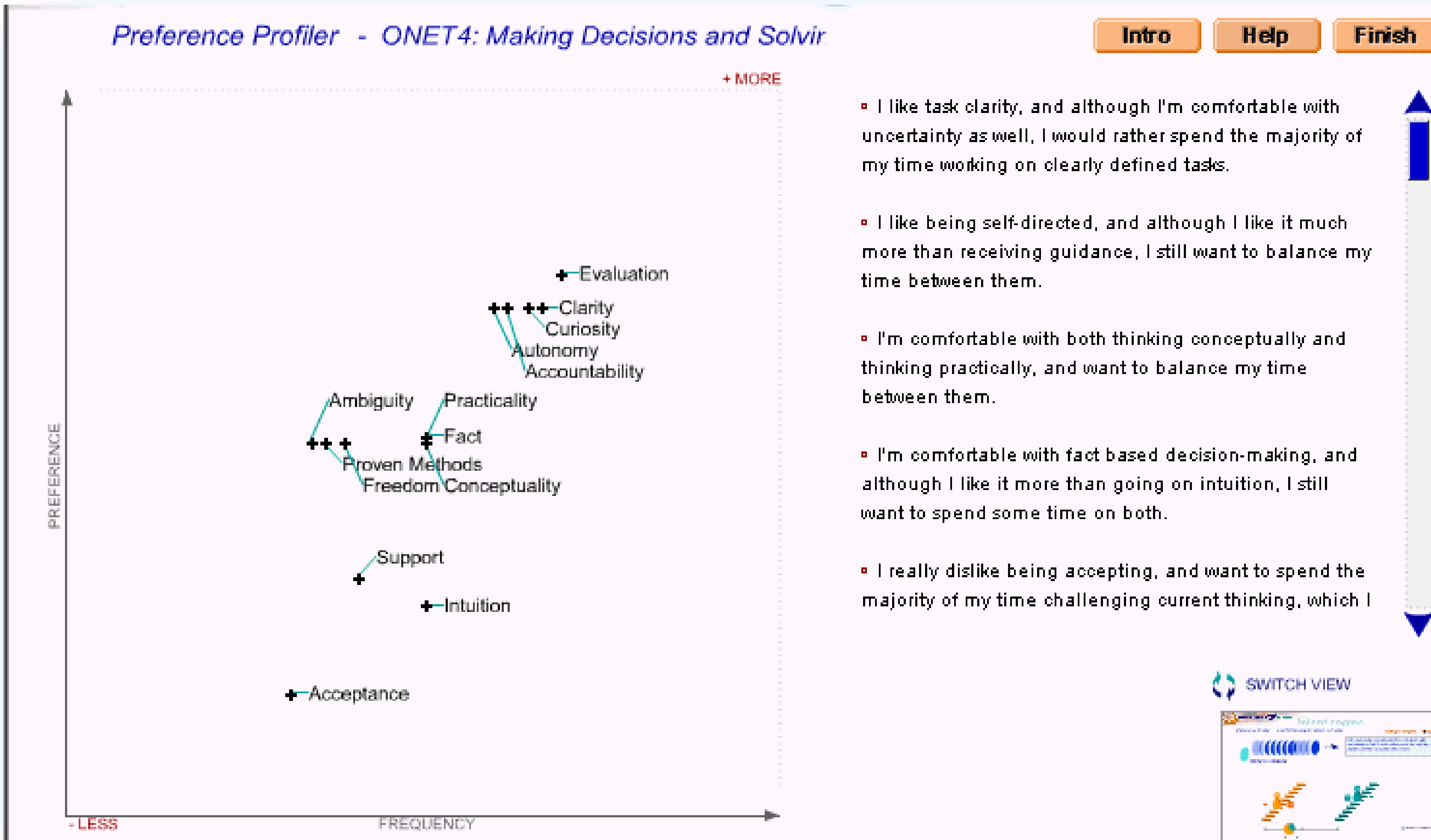


The Ideal Profile

- Use a 2-D profiler like the Mariner7-Staffcv profiler to construct a desired profile (as was done for the 2-D Work Preference-to-**O*NET™** application)

*O*NET* Goal #4: Making Decisions and Solving Problems

Analyzing information and evaluating results to choose the best solution and solve problems.



The "Star Performer" Profile

Literally a single empirically generated profile, acquired from an individual designated as a “star performer”. The aim is to hire more people for a particular role who best match the “star performer” profile.

Likewise, for “team profiling”, each member with a particular team-role (shaper, evaluator, leader etc.) of a “star-performing” team is profiled, such that a new team might be assembled for a similar task, with members closely aligned as possible to that “winning team profile set”.

Homogenous Group Profile

Probably the most common form of empirical target profile construction. It requires isolating a group of individuals considered homogenous for the purposes of constructing a group-descriptive profile. Then, the profile is constructed according to the properties of the variables that are constituent of the profile (i.e. nominal, ordinal, equal-interval).

Homogenous Group Profile: Nominal Data

This might consist of a variable such as:

Experience in Market Sector:

a) Retail
b) Manufacturing
c) Service

The homogenous group might provide data like:

Sector	Frequency
Retail	12
Marketing	23
Service	5

Homogenous Group Profile: Nominal Data

Profile Target Value will be constructed from:

❶ Probability weight i.e use the relative frequency to adjust a profile variable …

Sector	Frequency	Relative Weight
Retail	12	12/40=0.3
Marketing	23	23/40=0.575
Service	5	5/40=0.125

Homogenous Group Profile: Nominal Data

❷ Use a "Production Rule" to generate a profile variable "score" …

If *Retail* then assign value 1
If *Marketing* then assign value 3
If *Service* then assign value 0

Homogenous Group Profile: Ordinal Data

This might consist of a variable such as:

360-Rated "Customer Focus" (previous year)

a) Outstanding
b) Good
c) Acceptable
d) Poor
e) Extremely Poor

Homogenous Group Profile: Ordinal Data

The homogenous group might provide data like:

Rating	Frequency
1=Outstanding	14
2=Good	20
3=Acceptable	4
4=Poor	2
5=Extremely Poor	0

Homogenous Group Profile: Ordinal Data

Profile Target Value will be constructed from:

❶ Probability weight i.e use the relative frequency to adjust a profile variable.
❷ Production Rule.

Rule	Output
If rating < 3 then	5
If rating=3 then	3
If rating > 3 then	0

Homogenous Group Profile: Equal-Interval Data

This might consist of a variable such as:

15FQ+ Personality Trait Score: Suspicious

Score = between 1 and 10, unit-stens

The homogenous group might provide data as:

Sten	Frequency
1	1
2	3
3	2
4	4
5	2
6	7
7	9
8	8
9	3
10	1

Homogenous Group Profile: Equal-Interval Data

Profile Target Value will be constructed from:

❶ Probability weight i.e use the relative frequency to adjust a profile variable.
❷ Production Rule.
❸ Mean/Median Score on the variable.

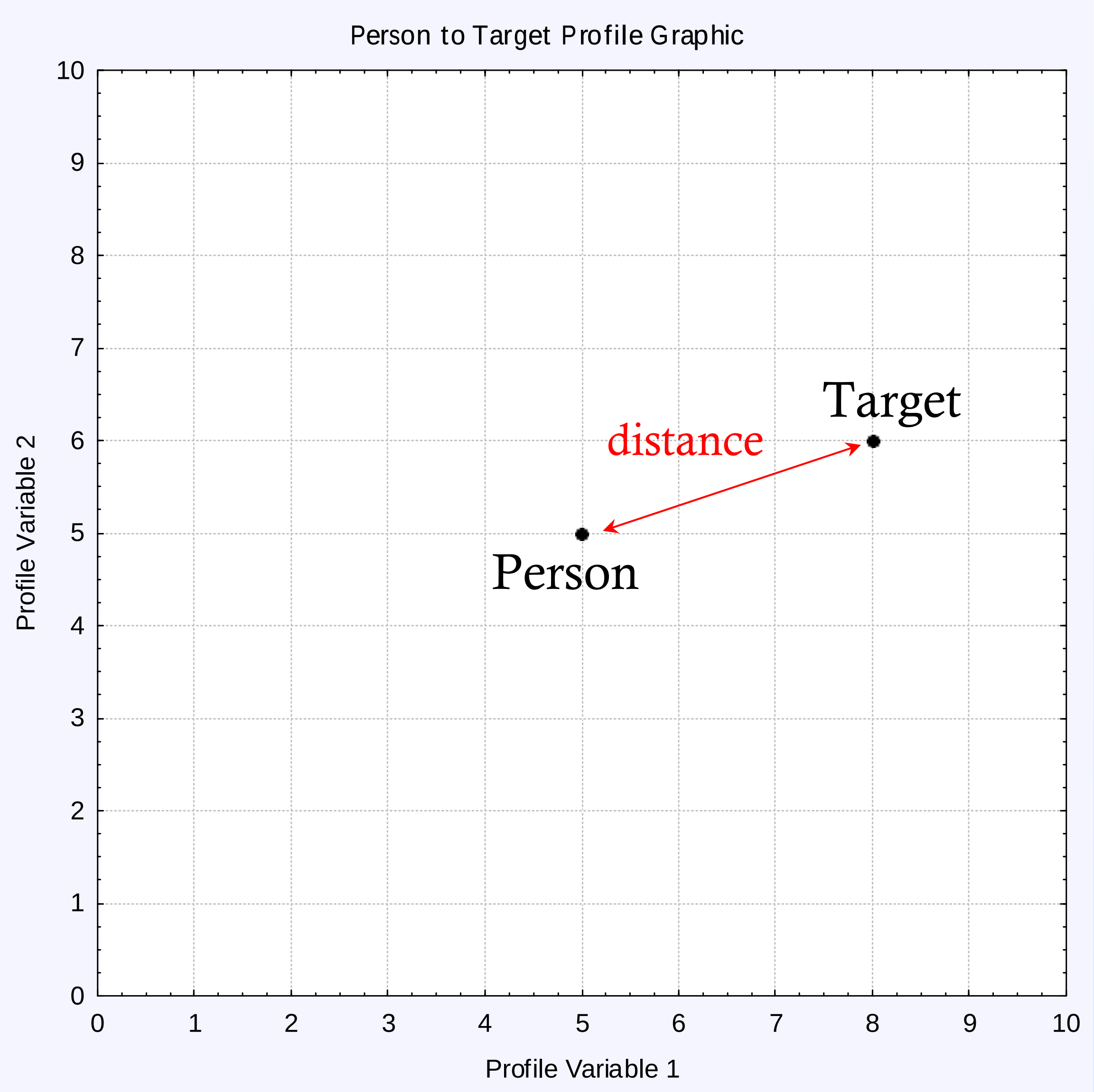
The Target Profile

Before we leave this Introduction to Profiling, let's consider the issue of Independent vs Correlated constituent components of a profile.

The Target Profile

Independent Profile Variables

Profile Var.	Target (t)	Person (p)
1	8	5
2	6	5



The Target Profile

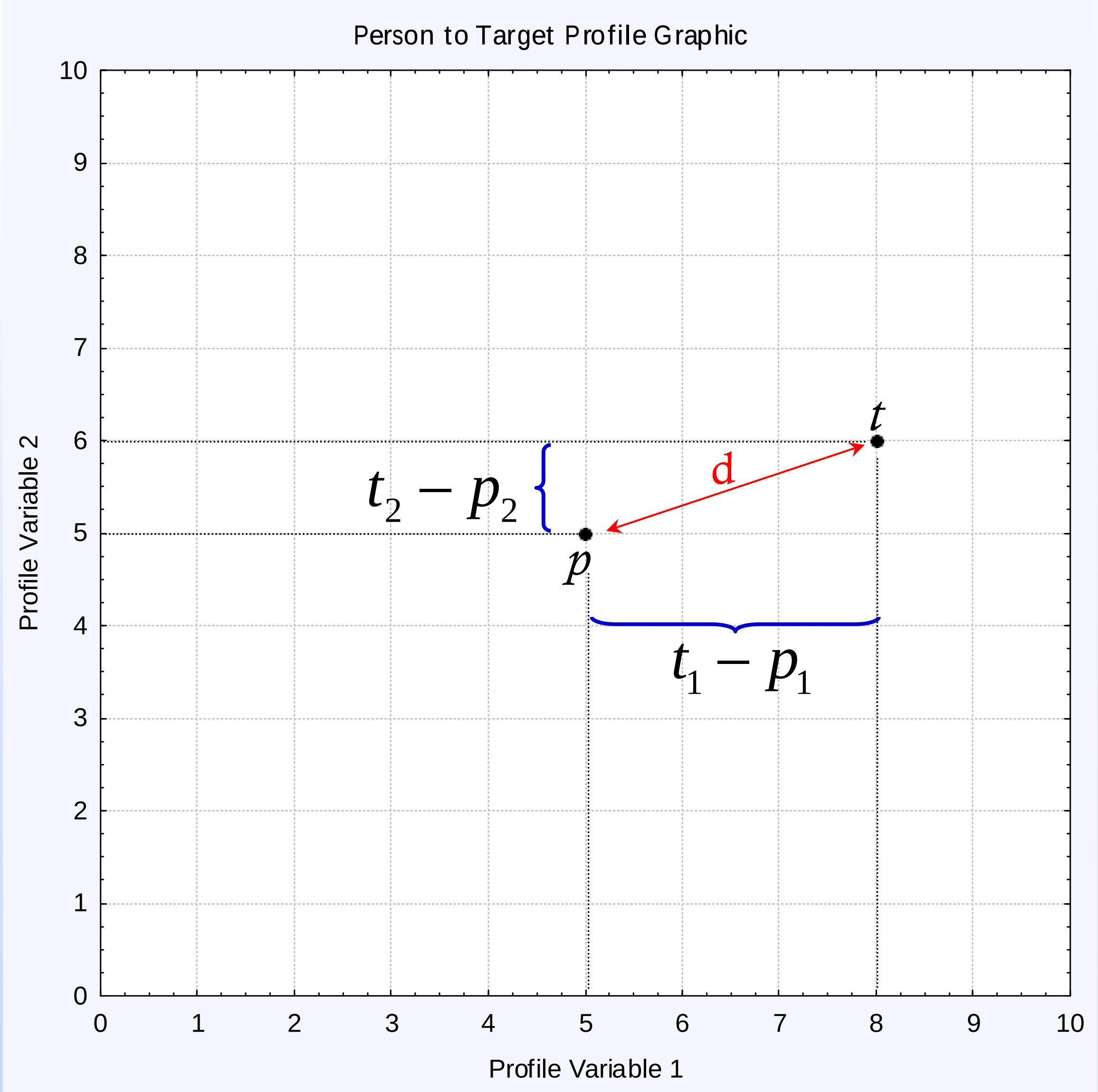
Independent Profile Variables

Simple Euclidean Distance

$$d = \sqrt{(t_1 - p_1)^2 + (t_2 - p_2)^2}$$

$$d = \sqrt{(8-5)^2 + (6-5)^2}$$

$$d = \sqrt{9+1} = 3.162$$

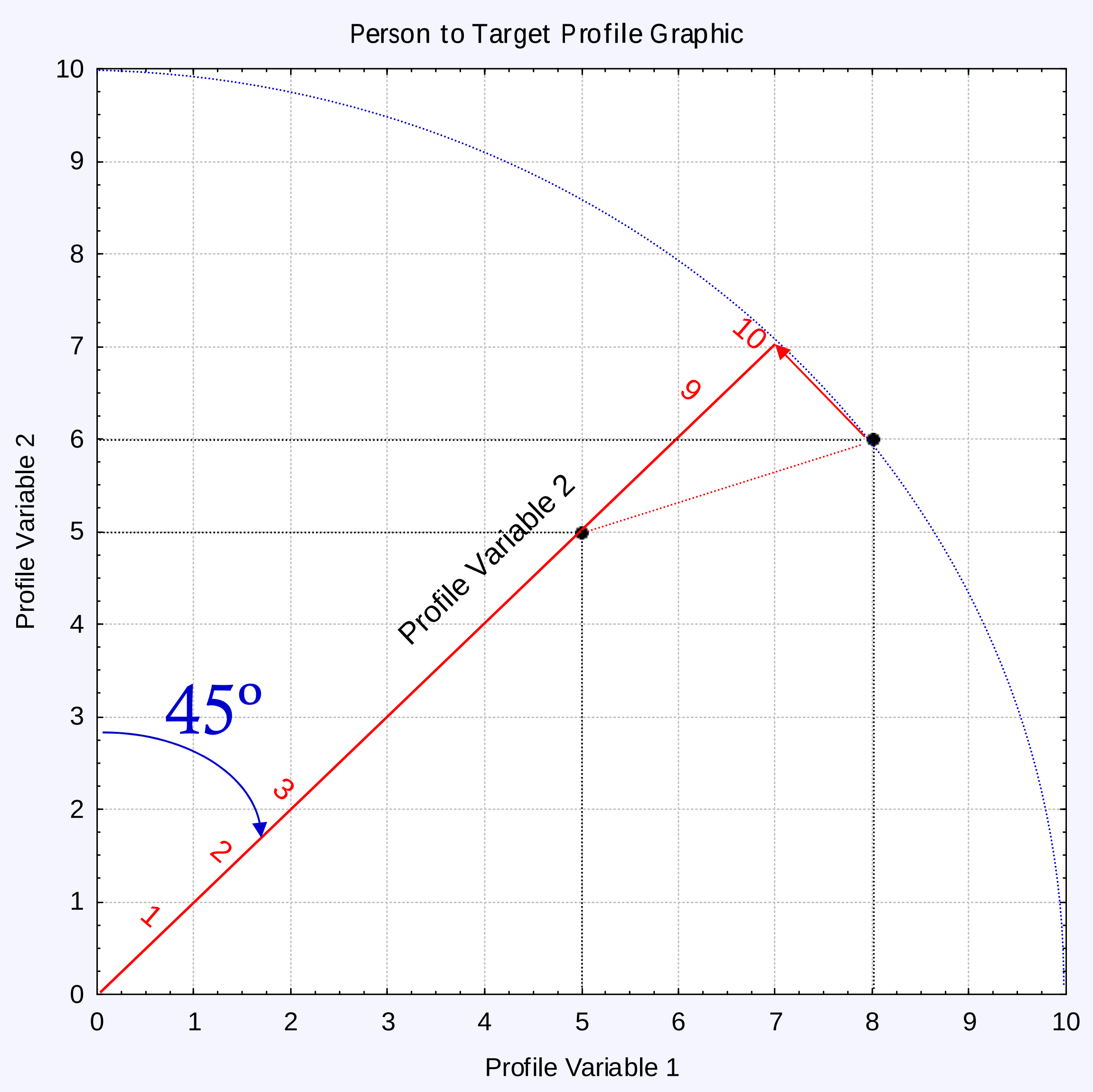


The Target Profile

Correlated Profile Variables

Profile Var.	Target (t)	Person (p)
1	8	5
2	6	5

profile variables 1 and 2 now correlate at 0.7071



Correlated Profile Variables

The usual statistic for expressing the "distance" of some multi-attribute profile from a target profile is **Mahalanobis Distance**. This uses the mean scores of the target attributes, along with the covariance matrix for the target attributes:

$$\Delta^2 = (x_i - \mu_i)^T \cdot \Sigma^{-1} \cdot (x_i - \mu_i)$$ which in our profiling notation is:

$$\Delta^2 = (p_i - t_i)^T \cdot \Sigma^{-1} \cdot (p_i - t_i)$$

where

$\Delta^2 =$ Mahalanobis distance

$\Sigma =$ covariance matrix of the target group data

$i =$ the i^{th} profile variable

The Target Profile Mahalanobis Distance

So, for our example, I need to generate a covariance matrix for some “target – homogenous group” data, where the means of each profile variable are 8 and 6 respectively, and where the correlation between variables is 0.7071.

The formula and process by which this is achieved is given in the document on the workshop CD entitled:

Likert Response Range and Correlation Attenuation.doc

The Target Profile Mahalanobis Distance

Profile Var.	Target (t)	Person (p)
1	8	5
2	6	5

profile variables 1 and 2 now set to correlate at 0.7071

$$\Delta^2 = (p_i - t_i)^T \cdot \Sigma^{-1} \cdot (p_i - t_i)$$

$$\text{dev} := \begin{bmatrix} (5-8) \\ (5-6) \end{bmatrix}$$

$$\Sigma := \begin{pmatrix} 1.056038 & 0.767689 \\ 0.767689 & 1.116319 \end{pmatrix}$$

$$\Delta\text{sq} := \text{dev}^T \cdot \Sigma^{-1} \cdot \text{dev}$$

$$\Delta\text{sq} = 11.0203$$

The Target Profile Mahalanobis Distance

However, the problem with Mahalanobis distance is that in order to use it, we need access to the covariance matrix for a target profile. This is rarely, if ever, possible to attain unless the target profile is constructed by a user from their own existing data, or the matrix is constructed by a test publisher as part of a "library" of target profiles.

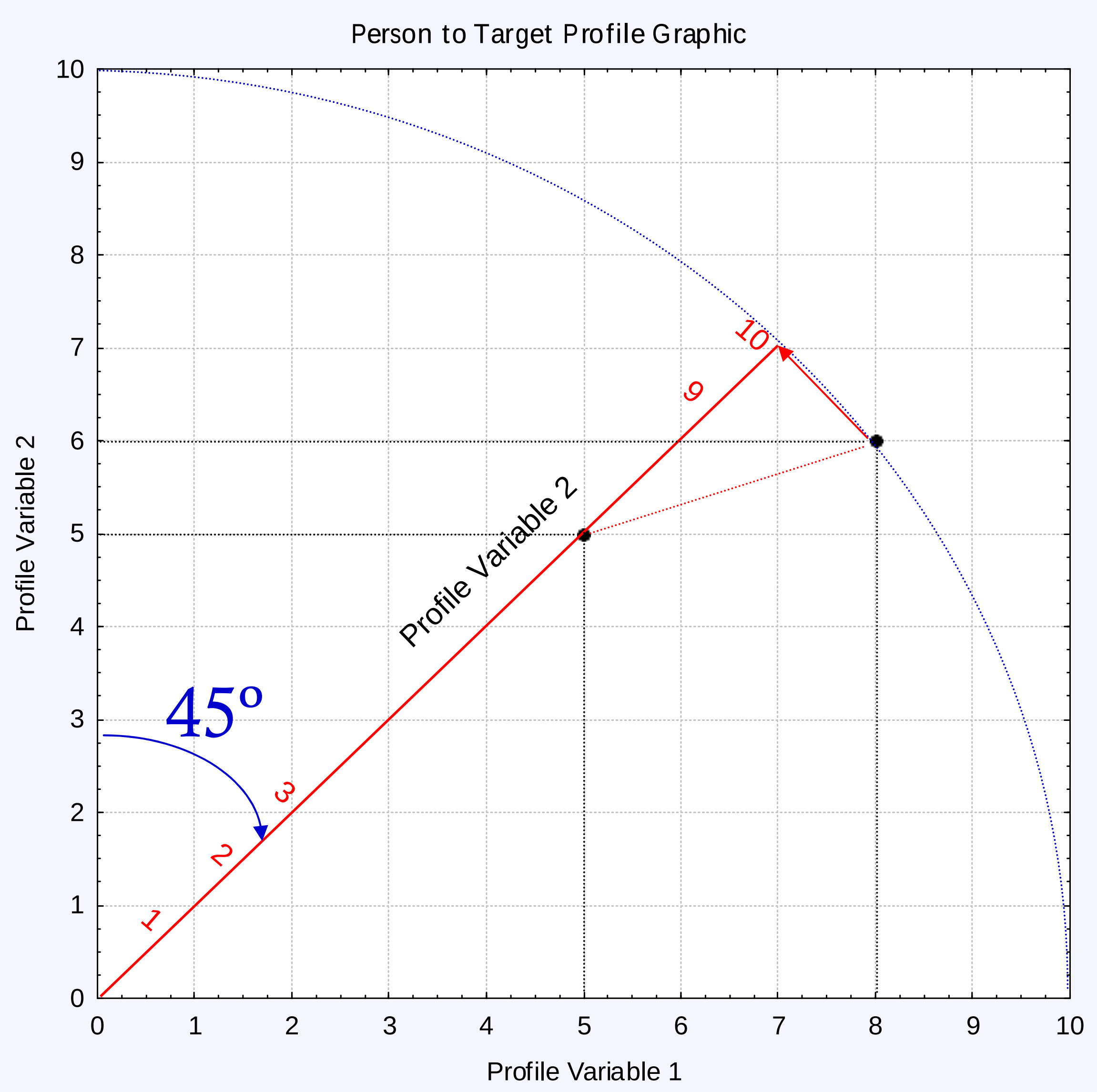
But, this latter idea is a non-starter. To assume that some generic profile will be suitable for ALL possible users is just not realistic. It may be possible to construct a "library" – but it requires replicated evidence across many user data-bases, allied to a unified set of criteria for which the target is being constructed.

The Target Profile

Correlated Profile Variables

So, we are back where we started. Is there any other way of computing distance/similarity using the data at hand? Maybe …

Profile Var.	Target (t)	Person (p)
1	8	5
2	6	5



Correlated Profile Variables

Given our orthogonal profile data, let's compute the new values for Profile Variable-2 as a result of the Profile Variable-2 axis being brought to within 45º of Profile Variable-1's axis.

Note that Pearson correlation is equal to the cosine of the angle between two variable vectors, hence a correlation of 0.7071 = arccos(0.7071) = 45º.

The Target Profile Correlation-Adjusted Distance

In order to reflect the effect of correlated "space", we have to adjust the values for profile variable 2, for both the Target and person profile accordingly. Using the standard formula for coordinate rotation …

$$t_2^* = r_t \cdot \sin(\varphi + \arccos(\rho))$$

with $\varphi = \arctan\left(\frac{t_2}{t_1}\right)$ and $r_t = \sqrt{t_1^2 + t_2^2}$ and $\rho =$ correlation coefficient

$$p_2^* = r_p \cdot \sin(\varphi + \arccos(\rho))$$

with $\varphi = \arctan\left(\frac{p_2}{p_1}\right)$ and $r_p = \sqrt{p_1^2 + p_2^2}$

The Target Profile Correlation-Adjusted Distance

Profile Var.	Target (t)	Person (p)
1	8	5
2	6	5

$$t_2^* = r_t \cdot \sin(\varphi + \arccos(\rho))$$

with $\varphi = \arctan\left(\frac{t_2}{t_1}\right)$ and $r_t = \sqrt{t_1^2 + t_2^2}$ and $\rho =$ correlation coefficient

$$t_2^* = \sqrt{8^2 + 6^2} \cdot \sin\left(\arctan\left(\frac{6}{8}\right) + \arccos(0.7071)\right)$$

$$t_2^* = 10 \cdot \sin(\arctan(0.75) + 45) = 9.899$$

The Target Profile Correlation-Adjusted Distance

Profile Var.	Target (t)	Person (p)
1	8	5
2	6	5

$$p_2^* = r_p \cdot \sin(\varphi + \arccos(\rho))$$

with $\varphi = \arctan\left(\frac{p_2}{p_1}\right)$ and $r_p = \sqrt{p_1^2 + p_2^2}$

$$p_2^* = \sqrt{5^2 + 5^2} \cdot \sin\left(\arctan\left(\frac{5}{5}\right) + \arccos(0.7071)\right)$$

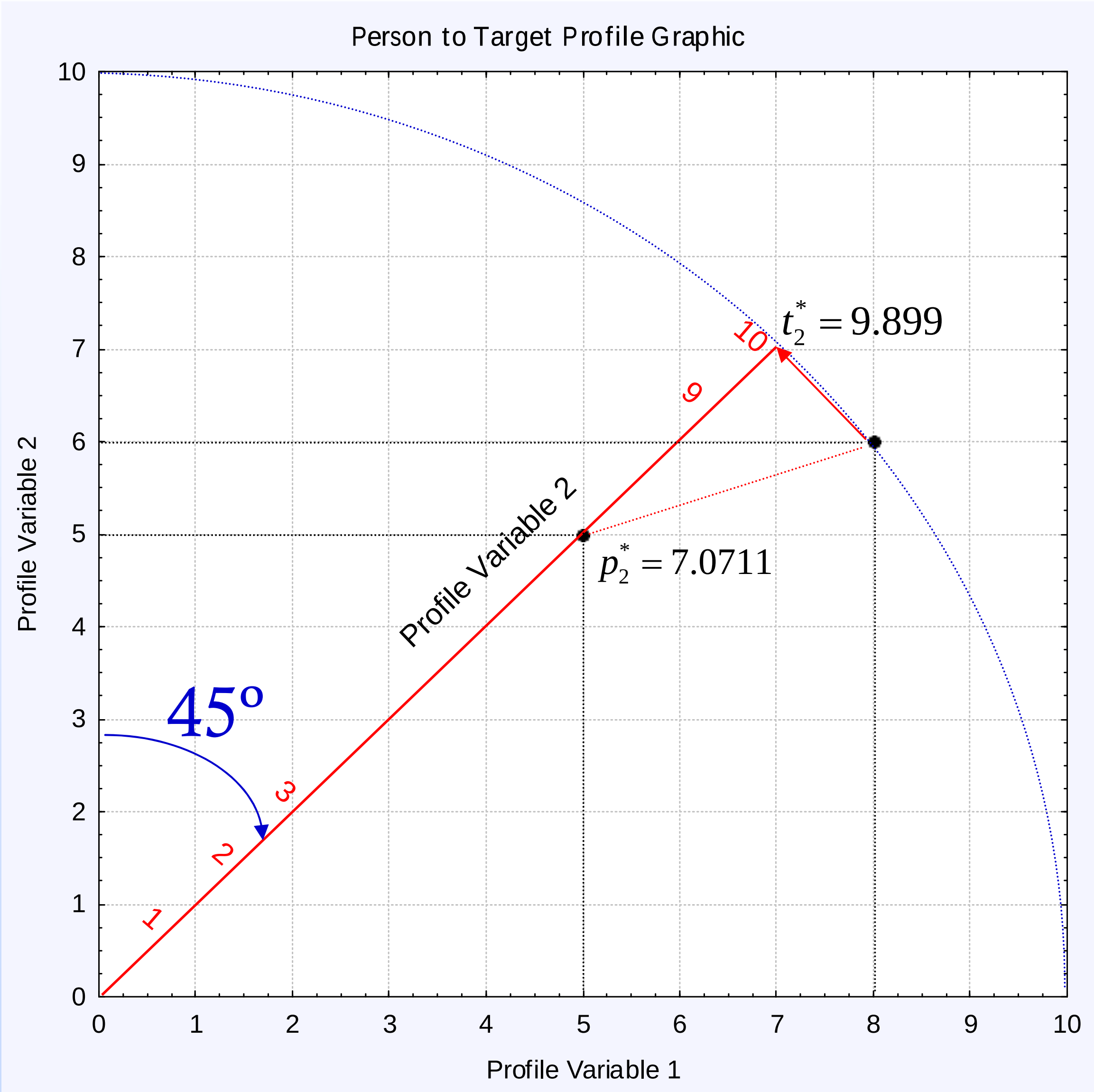
$$p_2^* = 7.0711 \cdot \sin(\arctan(1.0) + 45) = 7.0711$$

The Target Profile Correlation-Adjusted Distance

Correlated Profile Variables

Profile Var.	Target (t)	Person (p)
1	8	5
2	6	5

profile variables 1 and 2 correlate at 0.7071



The Target Profile Correlation-Adjusted Distance

Now, remember our initial scores (coordinates) were generated within an orthogonal framework. We have transformed them to represent what they would have been if they were observed within a space bounded by two oblique vectors (profile variables 1 and 2).

This allows a piece of "reverse logic" … if we know the correlation between two variables in advance, and observe two sets of scores on them, then we can compute what the scores would have been –IF- the scores had been acquired from two variables that were uncorrelated. In short, we "reverse-engineer" the coordinate calculations above.

The Target Profile Correlation-Adjusted Distance

Let's say we observed the following scores, knowing that our two profile variables correlated at 0.7071. We initially need to compute what the scores would have been given the variables were actually uncorrelated …

Profile Var.	Target (t)	Person (p)
1	8	5
2	10	7

The Target Profile Correlation-Adjusted Distance

The adjusted formulae are:

$$t_2 = \frac{t_2^* - t_1 \cdot \rho}{\sqrt{1-\rho^2}} \quad \text{with}$$

$t_2^* =$ observed target score from correlated profile variable 2

$t_1 =$ observed target score from correlated profile variable 2

and $\rho =$ correlation coefficient

$$p_2 = \frac{p_2^* - p_1 \cdot \rho}{\sqrt{1-\rho^2}} \quad \text{with}$$

$p_2^* =$ observed person score from correlated profile variable 2

$p_1 =$ observed person score from correlated profile variable 2

The Target Profile Correlation-Adjusted Distance

Profile Var.	Target (t)	Person (p)
1	8	5
2	10	7

profile variables 1 and 2 correlate at 0.7071

$$t_2 = \frac{t_2^* - t_1 \cdot \rho}{\sqrt{1-\rho^2}} \quad = \quad t_2 = \frac{10 - 8 \cdot 0.7071}{\sqrt{1-0.7071^2}} \quad = 6.142 \text{ (rounded to 6)}$$

$$p_2 = \frac{p_2^* - p_1 \cdot \rho}{\sqrt{1-\rho^2}} = p_2 = \frac{7 - 5 \cdot 0.7071}{\sqrt{1-0.7071^2}} \quad = 4.899 \text{ (rounded to 5)}$$

Which gives us back our original orthogonal scores - with some rounding

Profile Var.	Target (t)	Person (p)
1	8	5
2	6	5

The Target Profile Correlation-Adjusted Distance

Now we adjust the orthogonalised euclidean distance by a factor directly proportional to the correlation coefficient. This is done by considering the angular separation between the two profile variable vectors 1 and 2. With a correlation of 0.0, we have 90º separation. As the correlation increases, we effectively reduce the area within which distances may be measured. Basically, by moving from orthogonal to oblique space, there is less "distance-area" available in which to calculate score differences.

The Target Profile Correlation-Adjusted Distance

So, by calculating the arc-cosine of the correlation coefficient, this will give us the degree separation between the two profile vectors. Dividing this by 90º provides a relative weight (to the total available "distance space") that is used to re-weight the orthogonalised distance between the target and person profile scores (on two profile variables). The formula is:

$$d' = \left(\frac{\arccos(\rho)}{90}\right) \cdot \sqrt{\left(t_1 - p_1\right)^2 + \left(t_2 - p_2\right)^2}$$

$$d' = \left(\frac{45}{90}\right) \cdot \sqrt{(8-5)^2 + (6-5)^2} = 0.5 \cdot 3.162 = 1.581$$

The Target Profile Correlation-Adjusted Distance

For completeness, let's say we observed the original scores below, knowing that our two profile variables correlated at 0.7071.

Profile Var.	Target (t)	Person (p)
1	8	5
2	6	5

$$t_2 = \frac{t_2^* - t_1 \cdot \rho}{\sqrt{1-\rho^2}} \quad = \quad t_2 = \frac{6 - 8 \cdot 0.7071}{\sqrt{1 - 0.7071^2}} \quad = 0.485 \text{ (rounded to 0)}$$

$$p_2 = \frac{p_2^* - p_1 \cdot \rho}{\sqrt{1-\rho^2}} = p_2 = \frac{5 - 5 \cdot 0.7071}{\sqrt{1 - 0.7071^2}} \quad = 2.071 \text{ (rounded to 2)}$$

The Target Profile Correlation-Adjusted Distance

$$d' = \left(\frac{\arccos(\rho)}{90}\right) \cdot \sqrt{\left(t_1 - p_1\right)^2 + \left(t_2 - p_2\right)^2}$$

$$d' = \left(\frac{45}{90}\right) \cdot \sqrt{\left(8-5\right)^2 + \left(0-2\right)^2} = 0.5 \cdot 3.6056 = 1.8028$$

There are some nagging problems with this second approach to working with correlated profile attributes – but, as we shall see, the answer lies in empirically detailed simulations and coefficient calibrations.