The Semantic Space of Intellectual Humility

Markus Christen, University of Zurich, Switzerland

Mark Alfano, University of Oregon, USA

Brian Robinson, Michigan State University, USA

Overview

The Problem(s) under investigation:

- Exploring the disposition of Intellectual Humility
- Presenting a novel approach for data analysis & visualization

Methodology:

- Assumptions and the Thesaurus approach
- In-depth look at the visualization algorithm

Results:

- Some plausibility arguments for the methodology
- Three semantic maps on Intellectual Humility

Outlook:

- What we have learned
- Open questions

The Problem(s) under investigation

The Paradox of “Intellectual Humility”

Intellectual Humility (IH) is understood as “having a consciousness of the limits of one's knowledge, including sensitivity to circumstances in which one's native egocentrism is likely to function self-deceptively; sensitivity to bias, prejudice and limitations of one's viewpoint.”

Claiming to be (intellectual) humble, however, may involves a paradox of self-reference:

A rabbi falls asleep while praying. His students gather round, extolling his virtue in soft tones. During their panegyrics, he awakens, but pretends to sleep on. When a lull in their conversation arises, he opens his eyes wide, cocks his head to the side, and says, “And of my humility you say nothing?”

- **If you say you’re IH, then it’s doubtful that you are.**
- **If you say you’re not IH, then you *might* be.**

Measuring Intellectual Humility

In an ongoing project funded by the Thrive Center at Fuller Theological Seminary, we aim to develop explicit, implicit and behavioral measures for IH. We propose that IH is best conceptualized by exploring the nexus of vices it opposes:

Intellectual arrogance (IA), intellectual vanity, and related vices involve thinking too highly of your intellectual qualities and achievements, thinking too much about yourself, caring too much about what other people think about you – or else thinking too little of others, thinking too little about others, or caring too little about others' intellectual dispositions and accomplishments.

Intellectual Diffidence (ID), intellectual self-abnegation, and related vices involve thinking too little of yourself, thinking too much about your intellectual deficits, being too disposed to defer to others' opinions, and so on.

This requires a better understanding of the "semantic space" of IH, as all our measures and scales are language-based.

The Psycholexical approach

The basic idea of the psycholexical approach is that, all else being equal,

- a ***natural language*** is more likely to include a predicate for a property to the extent that the property is important to those who speak the language.
- the ***semantic structure*** of a language reflects to some extent the perceived structure of the phenomena described by the language.

We think that the psycholexical approach is promising in the investigation of IH because questionnaires are likely to be especially unreliable as measures of this construct due to the paradox of self-reference.

We propose to investigate the trait of IH psycholexically by comparing ‘intellectual humility’ with both its antonyms and synonyms..

Visualizing high-dimensional data

When analyzing a potentially large set of terms, we face a general problem of explorative data analysis: Visualization of high-dimensional data by means of a low-dimensional embedding.

Classical approaches for dimensionality reduction aim to represent the data structure on a linear subspace of the original data space:

- **PCA** performs a projection onto the axes with maximal data variance
- **MDS** finds a low-dimensional embedding that preserves inter-point distances

Problems:

- These methods perform poorly when applied to nonlinear data structures (which may be the case in our problem).
- Researchers are often faced with similarity or proximity data with correct ordering (which is the case), but potentially unreliable data values.

In this contribution, we use a novel approach that is able to deal with nonlinear structures in data space.

Methodology

What is a Thesaurus?

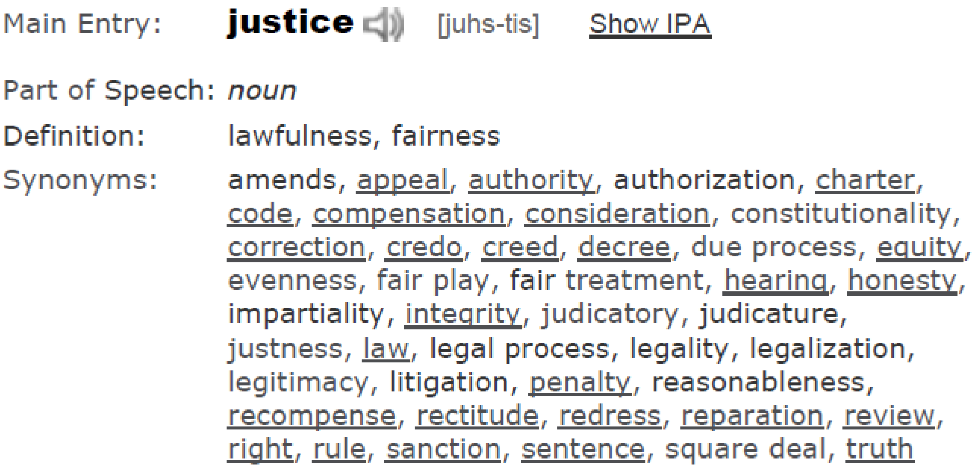
Our analysis is based on the assumption that the practice of language is precipitated in dictionaries, lexicons, and other wordbooks. Of particular interest is the thesaurus – a database of word similarities reflecting (written) language practice:

You are writing a text and you use the word "X" – but "X" does not quite express what you want to say: then you check a Thesaurus and you look for suggestions of words that have (somehow) a similar meaning.

Thus, a Thesaurus that emerged in decades of language use reflects word similarities based on actual use of the language.

- It's broader than "**WordNet**", that labels the semantic relations among words, whereas the groupings of words in a thesaurus does not follow any explicit pattern other than meaning similarity.
- Is less broad than determining the pure co-occurrence of words gained through **text mining** in a large document set.

Example: “Justice”



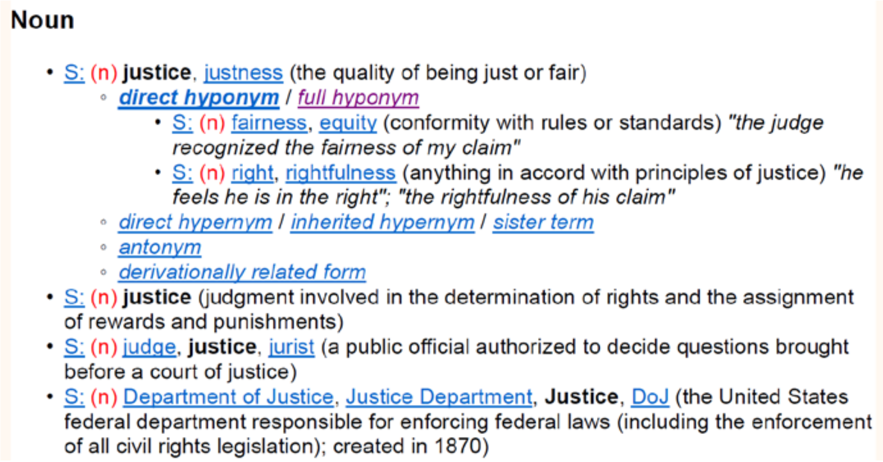
Main Entry: **justice** [juhs-tis] Show IPA

Part of Speech: *noun*

Definition: lawfulness, fairness

Synonyms: amends, appeal, authority, authorization, charter, code, compensation, consideration, constitutionality, correction, credo, creed, decree, due process, equity, evenness, fair play, fair treatment, hearing, honesty, impartiality, integrity, judicatory, judicature, justness, law, legal process, legality, legalization, legitimacy, litigation, penalty, reasonableness, recompense, rectitude, redress, reparation, review, right, rule, sanction, sentence, square deal, truth

Thesaurus entry



Noun

- S: (n) **justice**, justness (the quality of being just or fair)
  - ◦ ***direct hyponym*** / *full hyponym*
    - S: (n) fairness, equity (conformity with rules or standards) *"the judge recognized the fairness of my claim"*
    - S: (n) right, rightfulness (anything in accord with principles of justice) *"he feels he is in the right"; "the rightfulness of his claim"*
  - ◦ *direct hypernym* / *inherited hypernym* / *sister term*
  - ◦ *antonym*
  - ◦ *derivationally related form*
- S: (n) **justice** (judgment involved in the determination of rights and the assignment of rewards and punishments)
- S: (n) judge, **justice**, jurist (a public official authorized to decide questions brought before a court of justice)
- S: (n) Department of Justice, Justice Department, **Justice**, DoJ (the United States federal department responsible for enforcing federal laws (including the enforcement of all civil rights legislation); created in 1870)

WordNet entry

Prodecure (1)

1) We identified a large set of potential synonyms and antonyms for IH through literature review, internet search and consulting psychological scales that measure similar concepts (e.g.: H factor in HEXACO)

2) Four raters excluded terms that are obviously not expressing the concept of intellectual humility or its vices, leading to 52 synonyms and 69 antonyms of IH.

3) For each term *t* identified (both in noun and adjective form), we created its word-bag $T = \{t; t_{\text{syn1}}; t_{\text{syn2}}; t_{\text{syn3}}; \ldots ; t_{\text{syn}n}\}$, which is the set of all synonyms of the term in the database “Thesaurus.com”.

4) We calculated the pairwise similarity of all terms as the relative overlap of the associated word bags:

$$S(t^1, t^2) = \frac{|T^1 \cap T^2|}{\min\{|T^1|, |T^2|\}}$$

Prodecure (2)

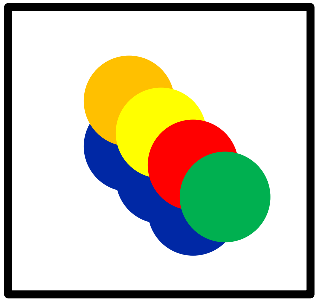
5) We checked for highly similar terms (S > 0.5) in order to reduce the number of terms for simplifying the analysis. We collapsed the word-bags of highly similar terms into a single word bag. In this way, we reduced the number of synonyms from 52 to 39 and the number of antonyms from 69 to 46. Then, the similarities were re-calculated.

6) The similarity measures obtained in this way were then used as inputs in a visualization algorithm called *superparamagnetic agent mapping SAM* (see next slides)

7) Finally, using a clustering paradigm that employs the same principles as SAM, we identified clusters on the map generated in step 6..

The Conceptual Idea of Superparamagnetic Agents

- The core idea of our approach is to translate the data into a set of agents.
- These agents ’construct’ a low-dimensional representation of the data in a self-organized way by moving according to laws of local spin interactions.



«Approach those who you like»

«Avoid those who you dislike»

Level 1 – Spin System (1)

We assume a given data set with *n* data points and its corresponding dissimilarity matrix with values $g_{ij} = g_{ji}$. Our method can be divided into two levels.

1. In the first level, each data item is represented by a Potts spin variable and the dissimilarity matrix is encoded in the spin couplings J_{ij} (k: k-nearest neighbors; a: average distance between neighbors):

$$J_{ij} = J_{ji} = \frac{1}{k} \exp\left(\frac{-g_{ij}^2}{2a^2} \right)$$

2. The spin system is treated in the formalism of the canonical ensemble, giving the probability for a certain spin configuration s_i as follows:

$$p(s) = \frac{1}{Z} \exp(-H(s)/T) \qquad H(s) = \sum\nolimits_{(i,j)} J_{ij}(1 - \delta_{s_i s_j})$$

Level 1 – Spin System (2)

3. By introducing a temperature-like parameter *T*, a cluster hierarchy can be generated. For smaller *T*, all spins tend to be in the same state. Upon an increase in *T*, large clusters break up into smaller clusters in a cascade of (pseudo-)phase transitions. For small *T*, spins that belong to data items of a noisy background can be filtered out as singletons that do not cluster (*M*: number of Monte Carlo Steps).

$$G_{ij} = \frac{1}{M} \sum_{t=1}^{M} \underbrace{\delta_{s_i^t s_j^t}}_{G_{ij}(t)}$$

The result the calculations of level 1 is a specific spin configuration. It serves as input for the calculation on levels 2 that moves the points representing data on the plain. After the calculations on level 2, a new spin configuration is calculated.

Level 2 – Agent System (1)

In the second level, each data item is represented by an agent in a 2-dimensional coordinate system and the agents move according to laws that are governed by the local interactions of the spin system. The algorithmic procedure is summarized as follow:

1) Choose a random distribution of agents in $\mathbb{R}^2$, a random spin configuration s^0 and set the temperatures $T=T_{min}$, T_{max} (both in the superparamagnetic phase) as well as ΔT (in dependence of M).

2) For T, calculate a new spin configuration s^{t+1} (Swendsen-Wang algorithm) and then the actual pair correlation $G_{ij}(t+1)$

3) Calculate the pairwise attraction / repulsion of agents and relocate them (see next page)

4) Repeat the procedure starting from step 2 until $T=T_{max}$.

Level 2 – Agent System (2)

- If $G_{ij}(t+1) = 1$ and $J_{ij} > 0$ then

$$\begin{aligned} \vec{x_i^{t+1}} &= \vec{x_i^t} + \alpha \cdot (\vec{x_j^t} - \vec{x_i^t}) \\ \vec{x_j^{t+1}} &= \vec{x_j^t} + \alpha \cdot (\vec{x_i^t} - \vec{x_j^t}) \end{aligned}$$

- else

$$\begin{aligned} \vec{x_i^{t+1}} &= \vec{x_i^t} + \beta \cdot e^{-d_{ij}^t} \cdot (\vec{x_i^t} - \vec{x_j^t}) \\ \vec{x_j^{t+1}} &= \vec{x_j^t} + \beta \cdot e^{-d_{ij}^t} \cdot (\vec{x_j^t} - \vec{x_i^t}) \end{aligned}$$

where $d_{ij}^t = |\vec{x_i^t} - \vec{x_j^t}|$.

Methodological remarks

(Optional) noise cleaning is performed by removing agents whose spins are in no clusters even for T_{min} (calculated in step 1).

Usually, the procedure is repeated for several temperatures T and then the mean location of the points is calculated.

The method does not offer unique solutions, which highlights the importance of the parameters involved: $0 < \alpha < 0{,}5$ controls attraction, $0 < \beta$ controls repulsion. Simulations show that α and β strongly determine the scaling of the final agent configuration:

- α mainly affects the intra-cluster distances.
- β mainly affects the inter-cluster distances.

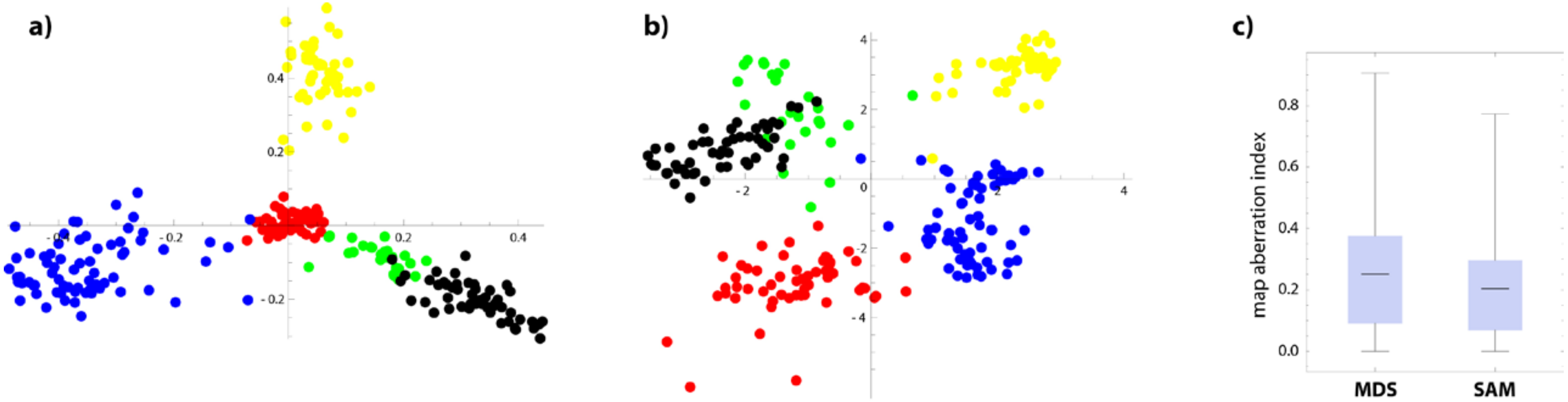
For the examples in this paper we used the values $\alpha = 0.1$ and $\beta = 0.01$ that have proven useful to balance inter- and intra-cluster distances.

Results

Plausibility arguments – Map quality (1)

We display data on an ongoing survey study that maps the similarity of scientific disciplines.

Map aberration index: We calculated for each item the sum of the absolutes of the normalized distance differences for each pair (original space vs. map space). The smaller the mean of this distribution (map aberration index), the better does the map preserve the topology of the original space.

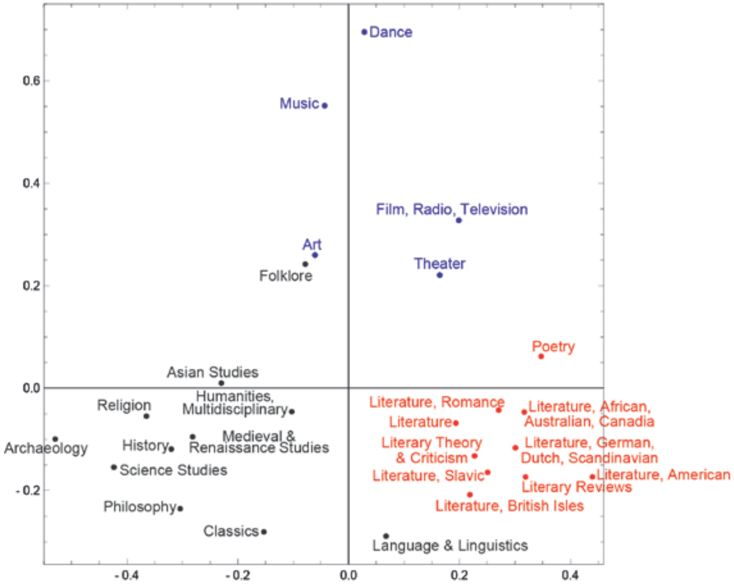


Blue: science; red: engineering; yellow: medicine; green: humanities; black: social science

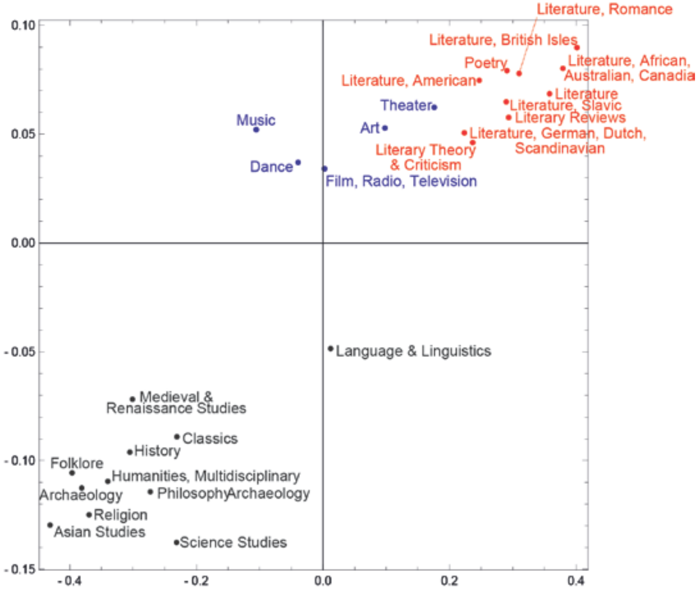
Plausibility arguments – Map quality (2)

When comparing the visualizations of the data for the group ”humanities”, for which most data was achieved in the survey, we find that SAM provides a more plausible representation:

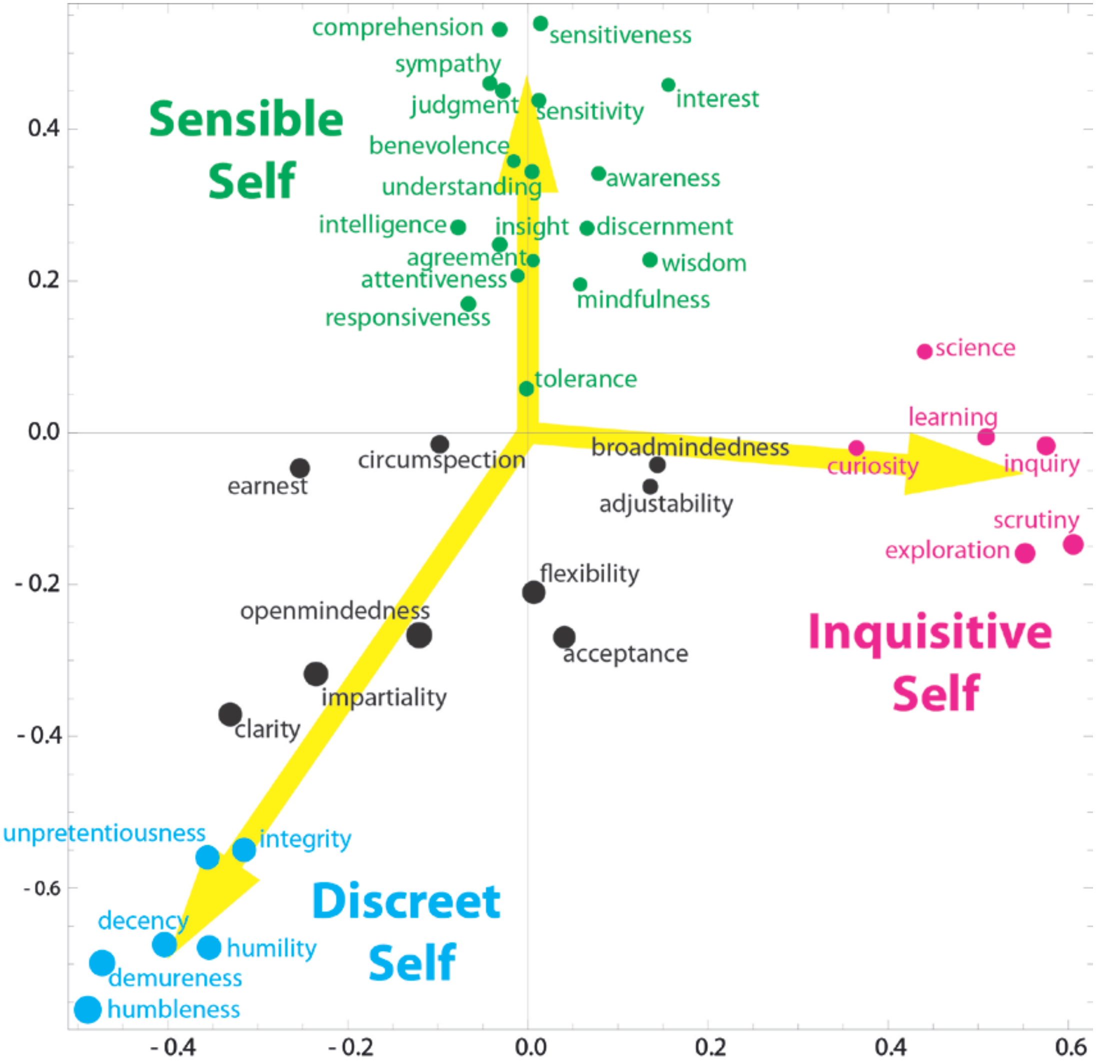
MDS



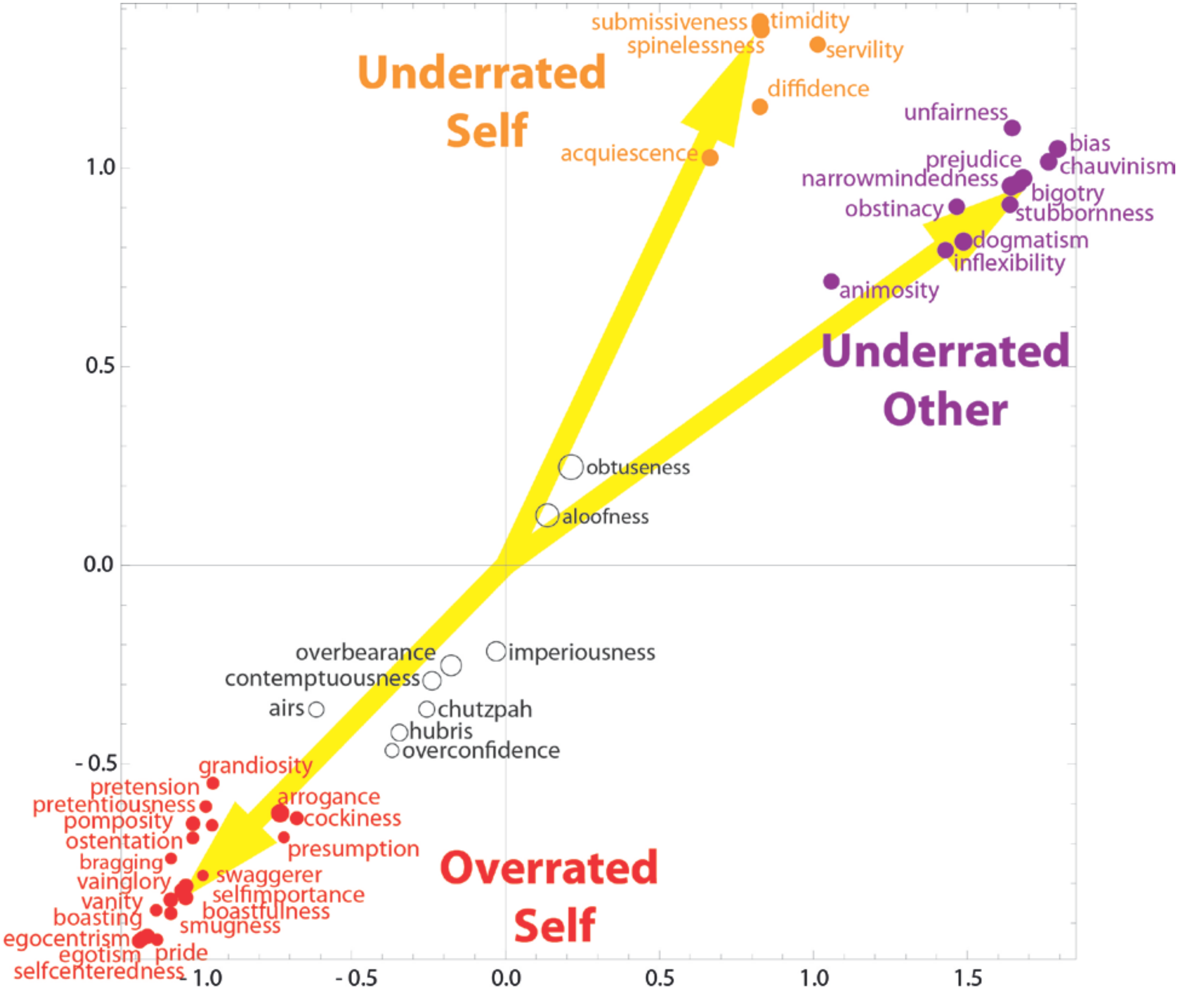
SAM



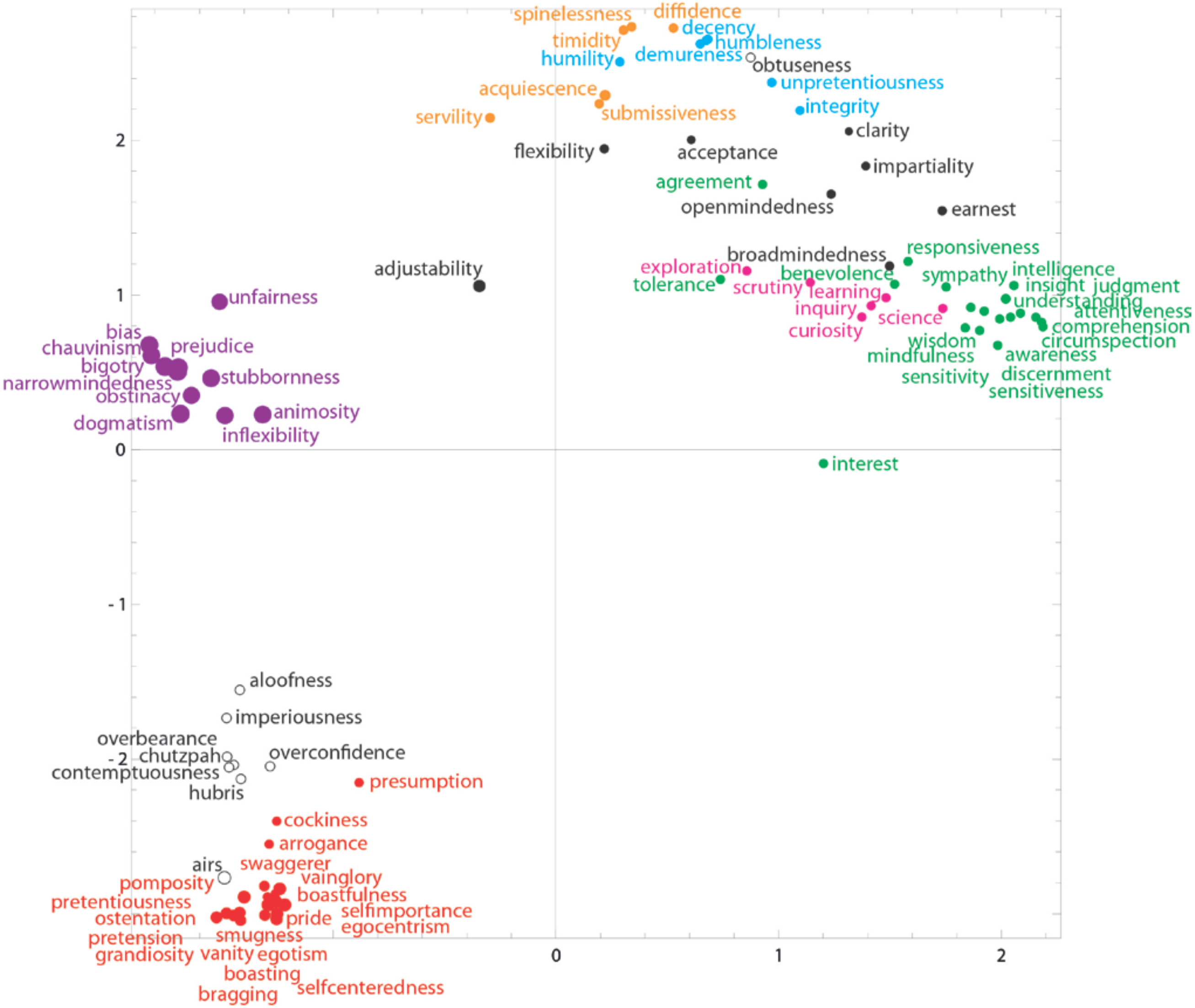
Results – IH Synonyms



Results –
IH Antonyms



Results – Synonyms & Antonyms



Outlook

Some key findings (regarding the IH study)

- The semantic understanding of intellectual humility (and its vices) has different facets that have to be taken into account when constructing scales and methodologies intending to measure IH

- One aspect of intellectual humility – the “discreet self” – has a semantic affinity to a IV vice (the “underrated self”). It might be that the discreet aspect of intellectual humility is essentially akin to underrating oneself. Alternatively, there might be two different traits picked out by these clusters – one a virtue and the other a vice – that are behaviorally similar enough that they are easily conflated. This may point to a potential problem of the Big Six personality inventory.

Open Questions – IH Study

We remind three shortcomings of the study

1) The nature of the data does not allow to decide, whether the similarities found only reflect purely pragmatic use of words within (American) English, or whether they actually represent relations between mental concepts.

2) The similarity measure used may misguide the analysis as we do not take into account differences in usage frequency of terms.

3) Our current methodology requires manual coding for term collapsing, as there is no biunique way of merging word bags. Although manual coding was feasible in our case, for larger term sets this approach may reach its limits.

Open Questions – Visualization Methodology

Although the heuristic superparamagnetic agents algorithm was successful in several applications, questions remain regarding the theoretical understanding:

- How can we quantify the role of the parameters α and β?
- How can the theoretical connections to other methods such as nonmetric multidimensional scaling be elaborated?
- Can we also use the technique to determine the true dimensionality of higher-dimensional data structures?
- What other rules or clustering methods could be used instead of our heuristics to generate a low-dimensional representation?

Thank you!