Thesaurus-based value maps as an instrument for psychological research

Markus Christen, Darcia Narvaez, Carmen Tanner, Thomas Ott

University of Zurich, University of Notre Dame, Zurich University of Applied Sciences

Table of Contents

Poster:
MC Larme, Ellen Roof, Olivia Godby, Darcia Narvaez, Markus Christen:
A German-English comparison of Thesaurus-based value maps reveals underlying semantic structures of cultural moral differences

University of Zurich[UZH]
Institute of Biomedical Ethics
University of Notre Dame
Psychology Department

The idea of “mapping a value space”

What are maps?

Maps are abstract, usually 2-dimensional representations of real (or abstract) spaces that involve entities, a metrics (i.e. a notion of distance between the entities), and interrelations among entities. A few basic principles are:

- Creating maps always involve **choices** on what to show and what not to show depending on the function of the particular map.
- Creating a map always involves the solution of a **classification problem** that relate map-symbols with the entities of the space one wants to map (e.g., single trees versus woods).
- Distance relations on maps are **incomplete** with respect to distance relations in the real space and may need additional symbols such that the map is not misleading (e.g., structural contour lines).

Attempts to “map a value space”

A “map” of a value space would involve:

- **Entities** that can potentially be an object of (moral) concerns (~values: states/ goals that individuals or groups/institutions consider to be achievable)
- A **feasibly number** of entities, i.e. they should emerge out of a classification (e.g., the moral foundations of Haidt, currently 6 classes)
- A **distance notion** between the entities (that’s why the moral foundations theory does actually not impose a map).
- A **problem** the map should help to understand.

An alternative way to create such a map would be to define “value dimensions” that span up a space in which non-moral entities are placed (e.g., Inglehart–Welzel cultural map of the world; the entities are countries)

Schwartz value map

Probably the most known “value map” emerged out of the research of Schwartz (function of the map: understanding psychological compatibility of values).

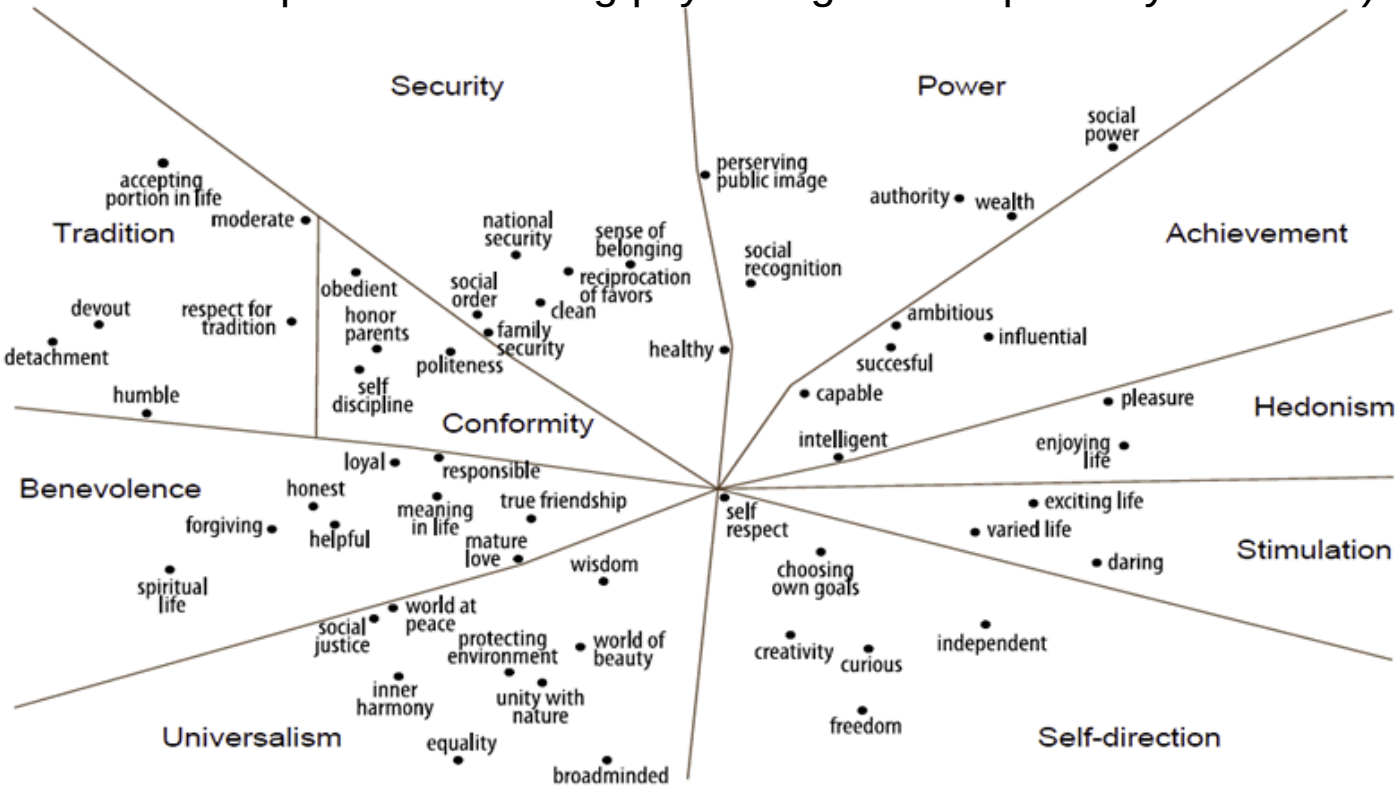


Fig. 2. The empirical structure of human values (Schwartz 1992).

University of Zurich[UZH]
Institute of Biomedical Ethics

University of Notre Dame
Psychology Department

Methodology

What is a Thesaurus?

A Thesaurus is 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

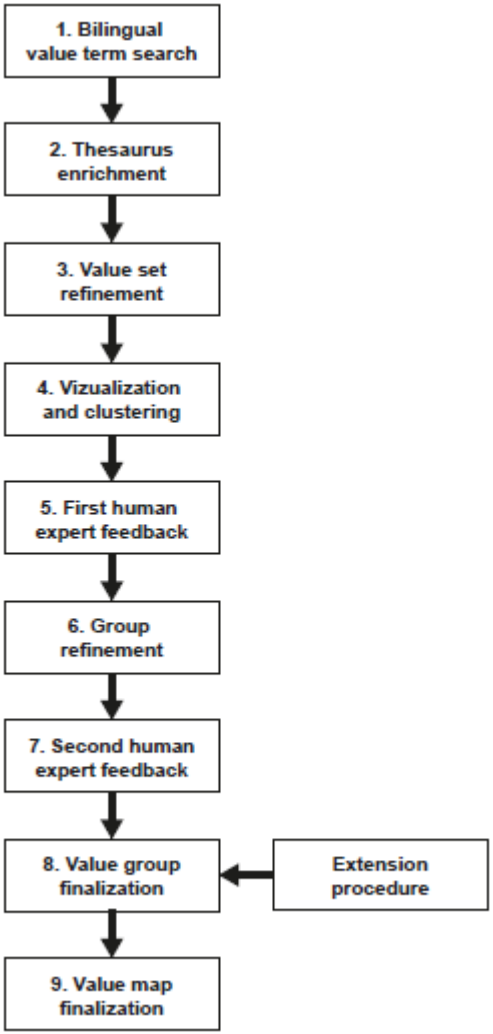
Synonyms: amends, appeal, authority, authorization, charter, code, compensation, consideration, constitutionality, correction, credo, creed, decree, due process, equity, evenness, fair play, fair treatment, hearing, honesty, impartiality, integrity, judicatory, judicature, justness, law, legal process, 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



Procedure (overview)

We used a “bottom-up” approach, i.e. we did not start with a classification, but we looked for value terms in all possible sources (literature, internet-lists, etc.) such that we had always a bilingual “match”.

Each term was then associated with a word-bag representing all the synonyms of the term (broadly understood). This imposes a distance relation (basically: word-bag-overlap).

Databases:
- Thesaurus.com (English)
- Woxikon.de (German)

The procedure also involves a mechanism to expand the map if new terms join the list.

Steps 1-3: Value term set generation

We used a broad definition of "value" as "**something that persons and/or institutions may consider to be desirable in a context that is likely not considered to be 'bad'**". For example, it includes terms like "aggression" as this may be desirable for some sports; but not terms like "cruelty" that may considered to be desirable for a criminal gang leader (to establish his position), because the context "criminal gang" lacks general approval.

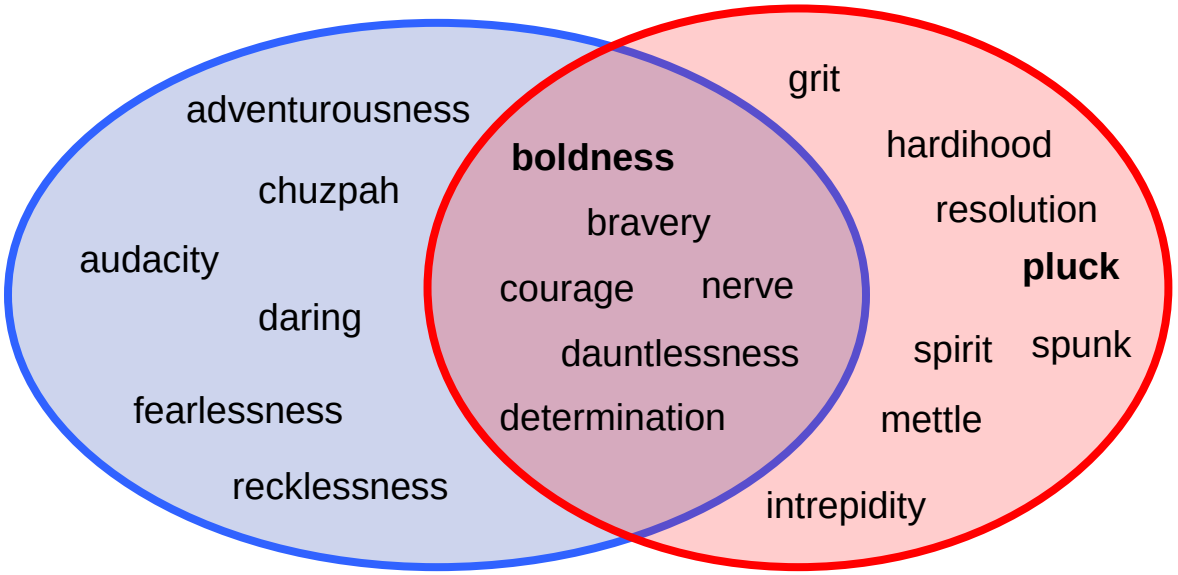
We made extended searches in dictionaries, value-lists, psychological & philosophical literature etc. to collect value terms both in English and German. We then also searched synonyms and refined our list, until we obtained a list of **460 value terms**.

In order to deal with **fuzziness in translation**, we ensured that, e.g., multiple German entries for translating an English term appeared as translations of similar English terms in the lists.

Step 4: Distance measure

For each value term, a **word-bag** has been created that contains all synonyms using thesaurus.com for English and woxikon.de for German.

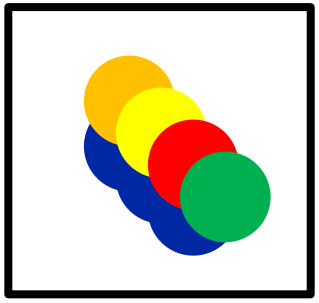
The **distance** between two terms is the relative overlap of the synonym sets (size of overlap divided by the size of the smaller word bag):



d = 6 / Min(12,14)
= 0.5

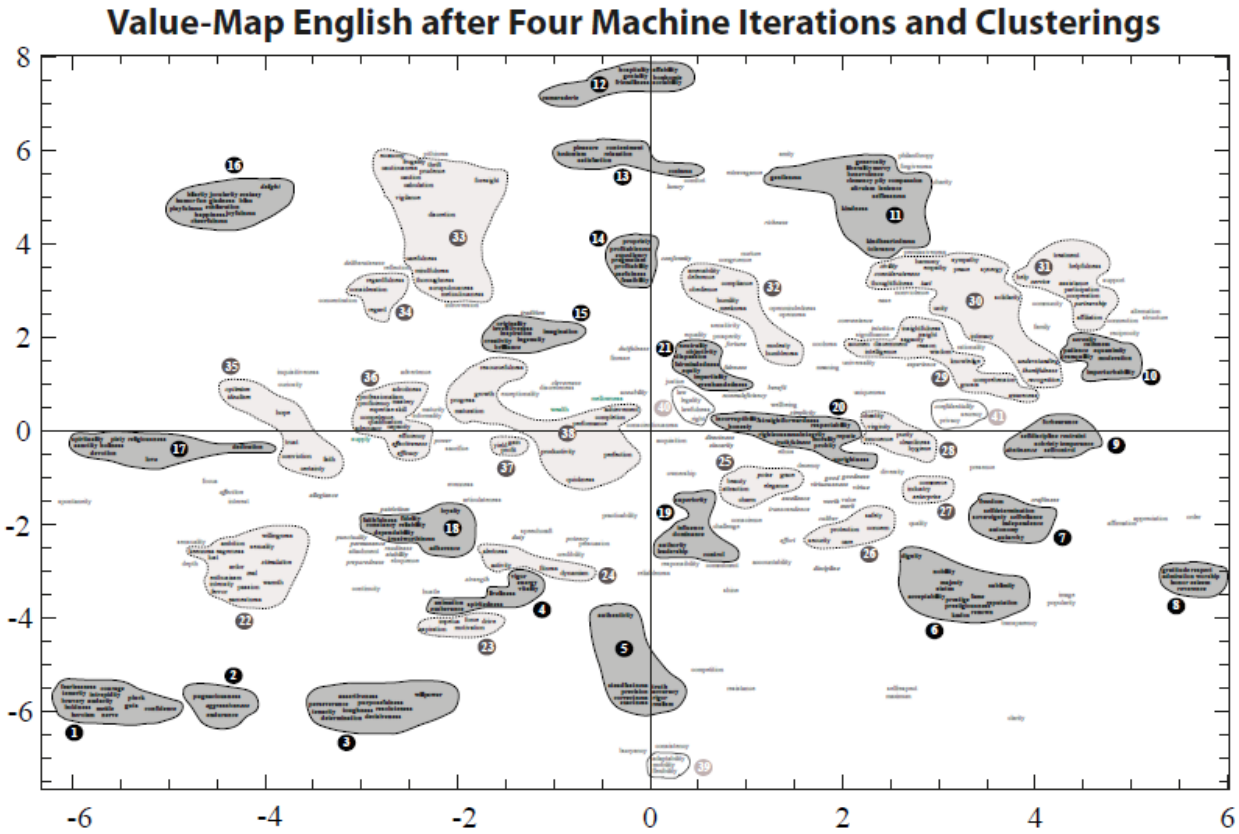
The idea of a superparamagnetic agent map

For visualization we used superparamagnetic agent maps (Ott & Christen 2011), a self-organization-based tool that preserves the topology of the high-dimensional space (including an iterative procedure: in the local environment of each point of the map: diminish only those distances of points that are also close in the real space)



Intermediate map (step 4)

After an iterative procedure to optimize the visualization and a clustering step (SSC) an intermediate map of all 460 values has been created.



Expert Evaluations (steps 5 / 7)

Human expertise has then been used (in total 9 experts per language; emerging from philosophy, psychology or English/German literature) to refine the classification:

Step 5: Five out of six experts had to agree with a classification. The groups then have been chosen such that inter-language match is preserved.

All remaining values have then been attributed to the “best” group in terms of synonym-overlap.

Step 7: All six experts per language had to agree in attributing a certain value to a group. In case of disagreement, best-match to the enriched groups has been calculated and the value has been rejected if a certain threshold has been missed.

University of Zurich[UZH]
Institute of Biomedical Ethics

University of Notre Dame
Psychology Department

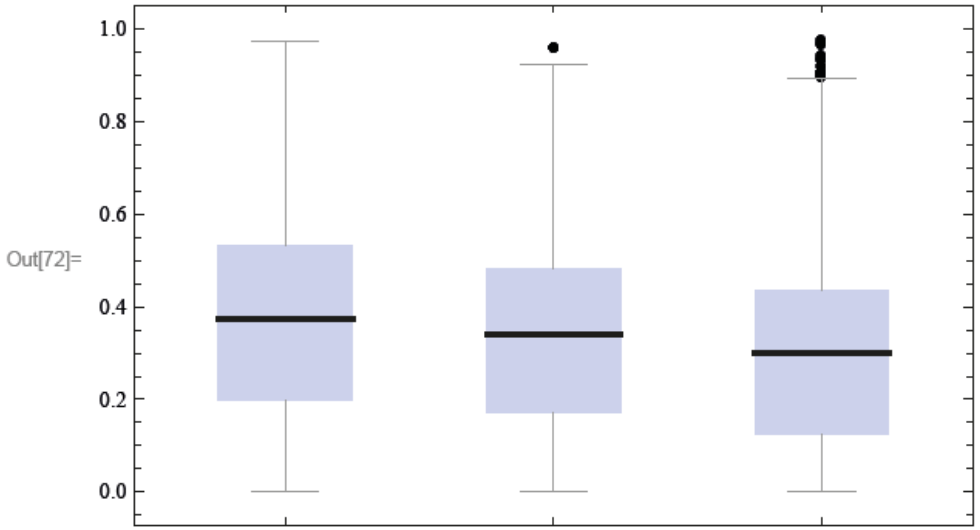
Results and result evaluation

Result: 78 groups have been identified (excerpt below)

Language-typical elements English	English values with German translations	rel. CS	rel. CS	German values with English translations	Language-typical elements German
	aggressiveness, pugnaciousness, toughness	0.86	0.29	**aggressivität**, härte, kampfeslust	herausforderung, widerstandsfähigkeit
benevolence, kindheartedness, kindness, thoughtfulness	**altruism**, charity, philanthropy, sacrifice, selflessness	1.12	0.86	**altruismus**, aufopferung, nächstenliebe, philantropie, selbstlosigkeit	gastfreundschaft, hilfsbereitschaft
realism	**authenticity**, clarity, truth, truthfulness	0.89	0.91	**authentizität**, klarheit, wahrhaftigkeit, wahrheit	gewissheit, glaubwürdigkeit, prägnanz
sovereignty	**authority**, influence, persuasion, power	0.28	0.60	**autorität**, einfluss, macht, überzeugungskraft	bestimmtheit, durchsetzungsver-mögen, kaltschnäuzigkeit, signifikanz

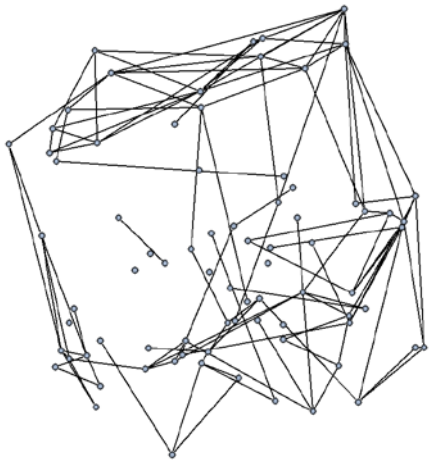
How to measure the quality of a map?

For each pair of points: calculate the aberration of their distance in the map to the distance in the real space (normalized with the longest distance in each case). The higher this value, the lower is the quality of the map in preserving the real distance relations.

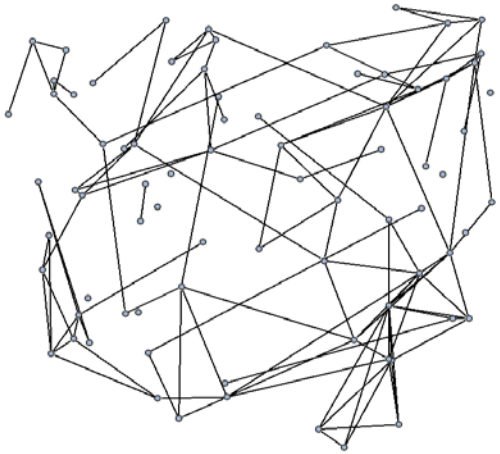


What does this mean for the map?

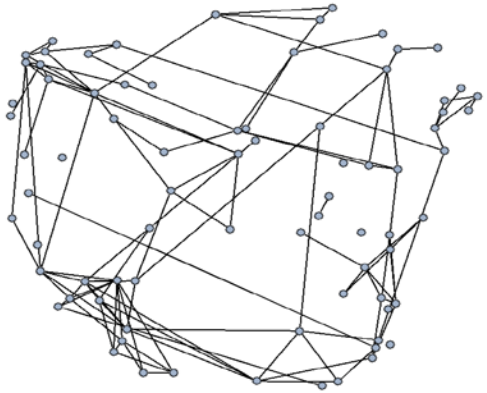
We show the maps (MDS vs. SAM1/2), where those points are connected whose distances in the original space is among the 5% closest.



MDS

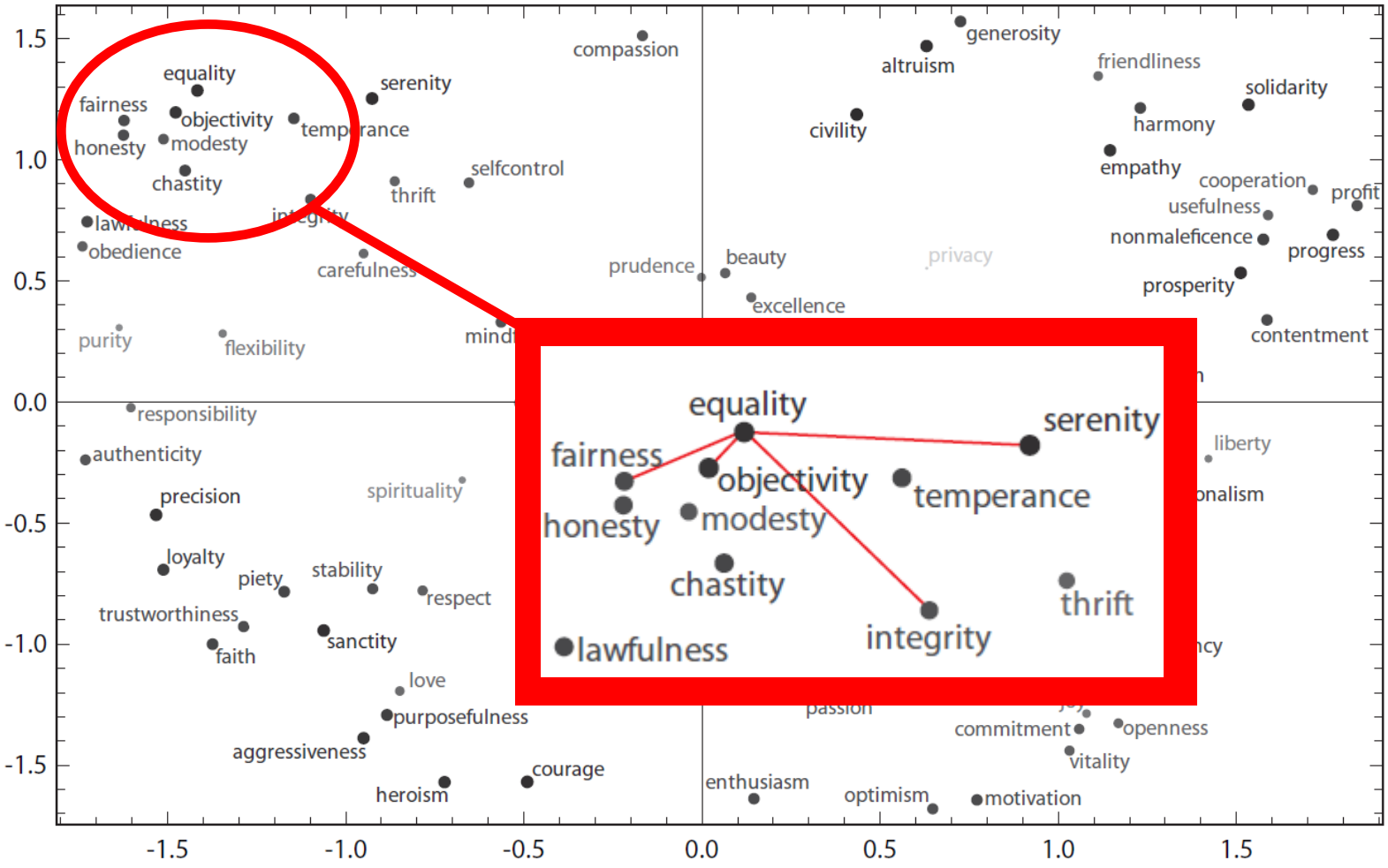


SAM1

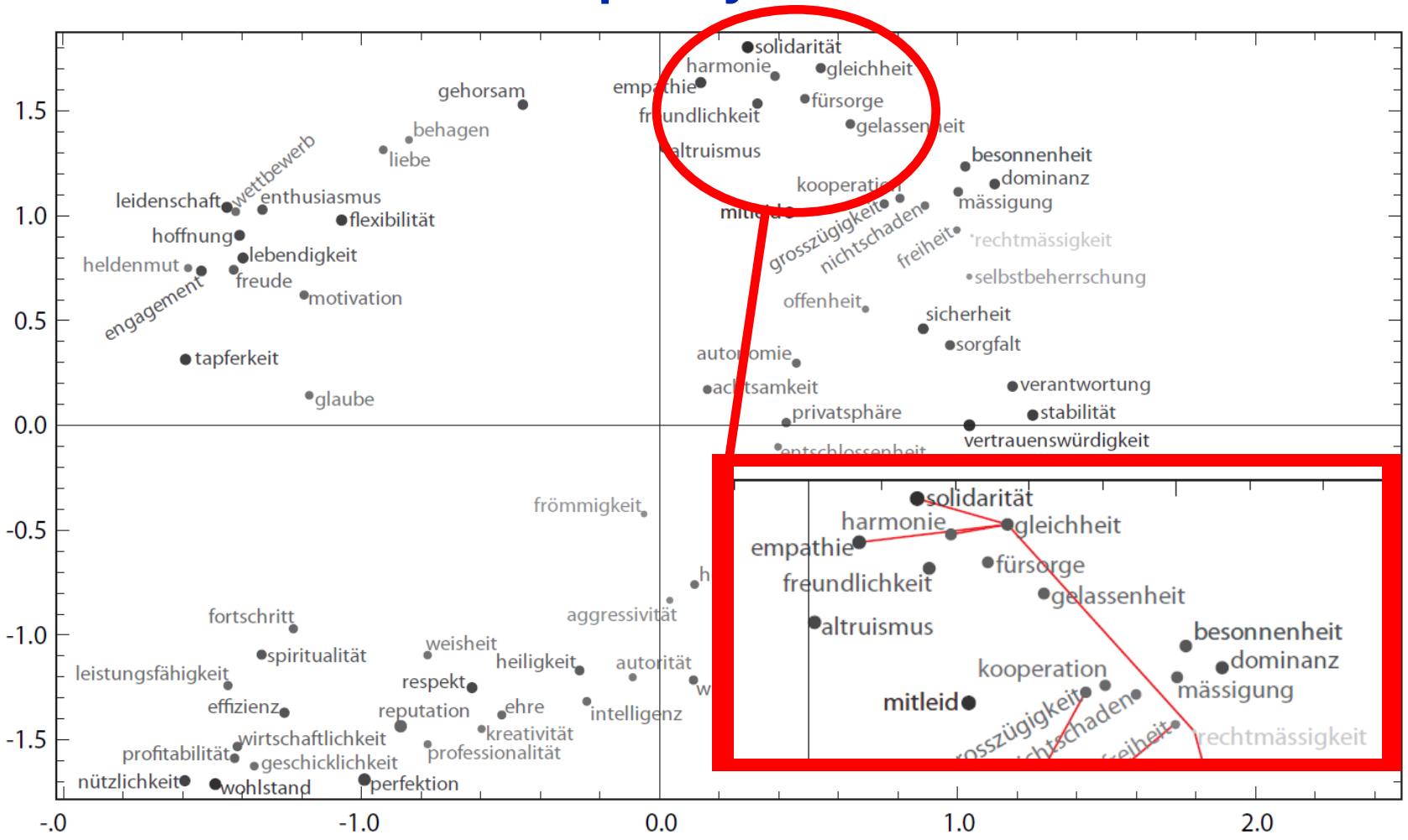


SAM2

Spotting cultural difference: equality/gleichheit: English



Cultural difference: equality / Gleichheit: German



Applications of the methodology

Our findings are of value both for methodological and research purposes:

1) Relying on “thesaurus similarity” as outlined in our study (i.e., the identified value groups) can be used to **optimize the translation of survey tools** across languages.
2) Understanding differences in semantic neighborhoods is **relevant for data mining** of social networks or digital communication, which becomes increasingly important for psychological research.
3) Maps resulting from such studies can be used as **exploratory tools** for identifying further differences with respect to the importance and semantic framing of values across cultures.

Next steps

Follow-up studies

Work in progress incudes:

1) Do people agree with the “within group similarity” of the value terms found in this analysis?

2) Does the thesaurus-based similarity across groups match with direct similarity rating?

3) Does closeness (on the map) imply higher accessibility (Higgins)?

4) To what extend are the value groups “morality laden” and is this dependent from the social context in which he values are used?

Validation of Thesaurus value groups by humans

Preliminary results of a study where participants (n=379; only USA) could delete “non-fit” synonyms from the value groups that emerged out of the Thesaurus study:

Median probability that a synonym was considered to be part of the group: 84.4% (mean: 83.7%)

Spearman Rank correlation between “group quality” assessed by humans and “cluster stability” of the Thesaurus groups: 0.26 (p=0.02).

This indicates that the Thesaurus results are in congruence to a direct similarity assessment of the value terms by humans.