Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 40 (2016) 59–74

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Using thesauruses as a heuristics for mapping values

Action editor: George Kampis

Markus Christen [a,*], Darcia Narvaez [b], Carmen Tanner [c,d], Thomas Ott [e]

[a] University Research Priority Program Ethics, University of Zurich, Zollikerstrasse 117, 8008 Zurich, Switzerland
[b] Psychology Department, University of Notre Dame, 100 Haggar Hall, Notre Dame, IN 46556, USA
[c] Department of Banking and Finance, University of Zurich, Plattenstrasse 32, 8032 Zurich, Switzerland
[d] Leadership Excellence Institute, Zeppelin University, Fallenbrunnen 3, 88045 Friedrichshafen, Germany
[e] Institute of Applied Simulation, Zurich University of Applied Sciences, Einsiedlerstrasse 31a, 8820 Wädenswil, Switzerland

Received 10 November 2015; received in revised form 10 February 2016; accepted 18 February 2016
Available online 19 March 2016

Abstract

Value differences across cultures or social groups are usually framed in terms of different emphases a particular group puts on specific values. For example, Western cultures typically prioritize values like autonomy and freedom, whereas East-Asian cultures put more emphasis on harmony and community. We present an alternative approach for investigating such cultural differences based on thesaurus databases that reflect the use of value terms in everyday language. We present a methodology that integrates empirical value research with linguistics and novel computer visualization tools to map and visualize value spaces. The maps outline variations in the semantic neighborhood of value terms. Based on 460 value terms both for US-English and German, we created for each language a map of 78 value classes that were further validated in two surveys. The use of such maps could inform research in three ways: first, by allowing for a controlled variability in the usage of value terms when generating vignettes; second, by indicating potential difficulties when translating value terms that display considerable differences in their semantic neighborhood; and third, as heuristics for better understanding value plurality.

Keywords: Cultural differences; Empirical value research; Semantics; Values; Value maps; Word similarity

1. Introduction

Values represent a prominent topic in many disciplines. Commonly understood as abstract and desirable standards, values are proposed to guide decision making and behavior. They are seen as crucial sources of conflicts within individuals or between groups and cultures (Huntington, 1993; Pearce & Littlejohn, 1997; Schwartz & Sagie, 2000). Another recurrent topic is whether there are universal values which should be uphold across cultures (Hare, 1954; Taylor, 1978) or whether values are relative and culturally determined (Quintelier & Fessler, 2012; Shweder, 1993). However, an important topic is also how to assess the meaning and structure of values. In psychology, many measures have been developed to assess the relative priority or importance of values (e.g., Rokeach, 1973; Schwartz, 1992) or to identify "innate" values (Haidt & Joseph, 2007). A prominent model of empirical value research is Schwartz's human value approach. Building on earlier models (e.g., Rokeach, 1973), Schwartz has developed two instruments to measure the importance of values, which have been widely tested and validated within and across cultures (e.g., Bilsky, Janik, & Schwartz, 2011; Schwartz, 1992, 2006). Based on a multidimensional scaling approach, he argues that values are organized in

* Corresponding author. Tel.: +41 44 634 8373; fax: +41 44 634 8507.
E-mail addresses: christen@ethik.uzh.ch (M. Christen), dnarvaez@nd.edu (D. Narvaez), carmen.tanner@bf.uzh.ch (C. Tanner), ottt@zhaw.ch (T. Ott).

http://dx.doi.org/10.1016/j.cogsys.2016.02.003
1389-0417/

a circular structure, reflecting ten distinct value domains with conflicts and congruity among values (Davidov, Schmidt, & Schwartz, 2008).[1]

Our research aligns with the general aim of classifying and mapping values, but uses a different approach. We consider this study as an example on how a specific way of knowledge representation (Thesaurus databases), machine classification and human expertise can interact for providing a solution for value classification. Specifically, this study is designed to investigate differences in the semantics of value terms in different natural languages. Our method is based on language usage reflected in linguistic reference books, offering a potentially ecologically valid approach of value identification, and furthermore involves a novel data analysis and visualization method that is based on self-organization. The general framework of our methodology relies on the tradition of psycho-lexical analysis that dates back to Francis Galton's *Measurement of Character* (1884). The basic idea 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. Furthermore, the psycho-lexical approach proposes that the semantic structure of a language reflects to some extent the perceived structure of the phenomena described by the language. In personality psychology, this insight was famously used by Allport and Odbert (1936) to create a semantic taxonomy of thousands of personality-relevant terms, which they argued represents how people conceive of personality. We aim for a similar type of analysis for value-related terms.

Our analysis is based on the assumption that the practice of language is precipitated in dictionaries, lexica, and other wordbooks. Of particular interest is the thesaurus – a language reference book or database organized to help its users find words related to a concept but having slightly different shades of meaning or connotation. Thesaurus dictionaries have a long tradition, starting in the 17th century and cumulated in famous books like Roget's "Thesaurus", published in 1852 (Hüllen, 2004). Thesauruses reflect what people in their daily use of language – in particular when writing text – consider semantically similar to a given term. They can be understood as expressions of "practical synonymy", which involves employing the principle of synonymy for semanticizing lexemes, i.e., basic units of lexical meaning that exists regardless of the number of inflectional endings it may have or the number of words it may contain (Hüllen, 2004).

Certainly, there is a rich theoretical tradition regarding the notion of synonymy in linguistics, philosophy of language and other fields. In a strict understanding, synonymy refers to the fact that there may be several different words for expressing exactly the same meaning – an understanding that is difficult to uphold, as Quine (1951) has observed. Within the emerging field of semantics, various notions of synonymy like semantic fields (Trier, 1931), the structuralist investigations of Harris (1973) or the pragmatic suggestion of Jones (1986) have been developed. In addition, sophisticated databases like, e.g., *WordNet* (see http://wordnet.princeton.edu/) that labels the semantic relations among words have been developed. However, today's thesaurus databases list synonyms in a broad sense, i.e., they employ some notion of "meaning similarity". The major aim of a thesaurus is not to find a replacement Y of a certain term X what has exactly the same meaning of X, but rather to find a term Y that has a slightly different facet in meaning for better expressing what the writer actually wants to express. Thus, a thesaurus is broader in capturing word relationships than synonymy in a strict sense – but it is still more specific than the mere co-occurrence statistics of terms in texts.

In what follows (Section 2), we will describe our methodology that involves both machine classification and expert opinion, while integrating classification algorithms and visualization tools. We start with a broad sample of value terms in two languages (460 terms each) that are then grouped into broader value groups using a two-step iteration procedure (see below, Fig. 1).

In Section 3 (Theory & Calculation), we present the application of a novel visualization tool such that experts can better deal with the high-dimensional data spaces that result from the large number of terms in our analysis

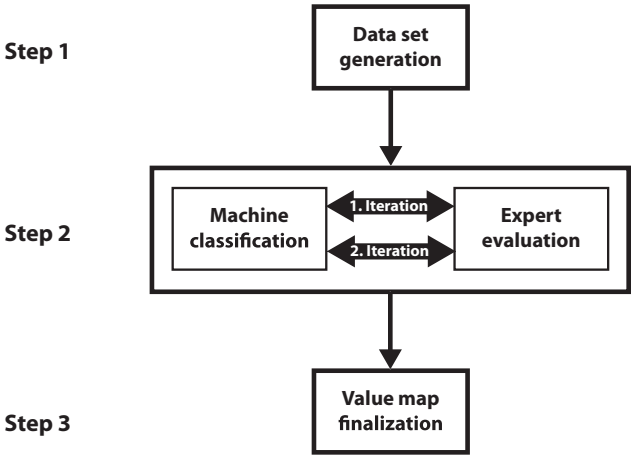

Fig. 1. Thesaurus value map generation procedure.

[1] The values theory of Schwartz defines ten broad value domains according to the motivation that underlies each of them: Self-Direction, Stimulation, Hedonism, Achievement, Power, Security, Conformity, Tradition; Benevolence, and Universalism. Those are considered to be universal because they are grounded in one or more of three universal requirements of human existence: needs of individuals as biological organisms, requisites of coordinated social interaction, and survival and welfare needs of groups. Schwartz proposes a circular arrangement of the values that represents a motivational continuum: The closer any two values in either direction around the circle, the more similar their underlying motivations; the more distant, the more antagonistic their motivations. For example, Achievement and Benevolence are opposite, meaning that former relies on self-centered satisfaction, whereas latter on devotion for peers. Conformity and Tradition are neighbors because they refer to the subordination of the self in favor of socially imposed expectations. A comprehensible introduction is provided by Schwartz (2012).

(Ott, Eggel, & Christen, 2014). We believe that this visualization technique, called superparamagnetic agent maps (SAM), has several methodological advantages compared to traditional methods. Because it is essentially local (i.e., only neighboring data points interact), SAM can deal with nonlinear structures and its nonparametric characteristics makes it robust towards noise. In particular, SAM has a higher topological reliability in mapping a high-dimensional space on a plane compared to multidimensional scaling (which is used in the Schwartz value map). In Section 4, we will show how these maps can be used as a heuristic to understand how cultural differences in value meaning emerge in the local neighborhood on value maps. In addition, we present results of a survey study that aims to validate the composition of the value groups by asking participants to group values into a synonym set based on their own understandings. Finally, in Section 5, we outline the relevance of our methodology as well as potential shortcomings for empirical value research.

2. Material and methods

2.1. Creating the thesaurus value map

In the following, we employ a broad definition of what constitutes a value. By "value" we denote any term that individuals or institutions consider being a positive goal worthy of achievement. For example, profit is a positive orientation in business and beauty is a positive one in art. We also include terms that are traditionally called virtues (e.g., generosity or curiosity) as well as terms that refer to orientations that are positively laden only in specific contexts, but are more ambiguous with respect to their desirability in other contexts (e.g., aggressiveness is a positive orientation in military and some sports, but not in other contexts). We exclude values that might be considered positive in some contexts but are generally disapproved (e.g., brutality may be a positive orientation for criminal gang leaders for upholding their status, but not so in general society).

Based on this understanding of values, we used the following procedure (summarized in Fig. 1) to generate two value maps (English and German) that contain 78 translated synonymous value groups each. In total, these groups include 440 English and 449 German value terms. Our method can be sub-divided in three steps that are discussed below.

2.1.1. Step 1 – Data set generation

Our methodology is inspired by a bottom-up approach. We aim to classify values based on a very large sample of terms instead of predefining a low number of values and analyzing their interrelations. We created our set of value terms in three phases:

(1) We started with an exhaustive search for both English and German value terms (nouns) by reviewing value research and terms in philosophical and psychological literature as well as on the internet. We paid particular attention to include the vocabularies of relevant value studies (Haidt & Joseph, 2007; Rokeach, 1973; Schwartz, 1992; Shweder, 1993), the classic virtues (outlined in the work of Plato, Thomas Aquinas, and others), as well as the vocabularies of desirable values listed in various websites (e.g., of organizations). For each term found in either language, we also included its translation in the other language using the *LEO* online dictionary, one of the largest and most popular German–English dictionaries (http://dict.leo.org).[2] A total of 448 value terms per language has been identified in this way.

(2) We identified for each value term the set of its synonyms. For the English synonyms, we used the largest English thesaurus provider *Dictionary.com* that contains more than one million words (http://thesaurus.com/). For the German synonyms, we used the largest German thesaurus provider *Woxikon* including more than 200,000 words (http://synonyme.woxikon.de/). All value terms were represented by nouns. In the rare cases where there was no thesaurus entry for a noun identified in phase one, we used the thesaurus result of the related adjective that was then nominalized. We were tolerant in accepting synonyms for a given word, only clear homonyms and slang expressions consisting of more than one word were not included into the synonym set of a specific term. In that way, we associate each identified term t with a word bag, which is the set of synonyms listed for that term in the databases. The *semantic constellation* of a term t is thus an ordered pair $(t, \{t, t_{syn1}, t_{syn2}, t_{syn3}, \ldots, t_{syn\,n}\})$, with the first element being t itself and the second element representing t's word bag (i.e. the set of synonyms of t including t itself).

(3) We cleaned our data set in the following way: We first checked whether there were value terms that appeared at least 10 times in the union of all word bags $\cup_{i=1}^{448}\{t^i, t^i_{syn1}, \ldots t^i_{syn\,n}\}$ but that were not yet present in the list generated in step 1. This was the case for 72 English and 59 German terms. Those terms

[2] There is ambiguity in translation (the "indeterminacy of translation thesis"; Quine, 1960, chap. 2) that shows up on a practical level. For example, the following German terms are proposed by *LEO* as translations of "generosity": Edelmut, Freizügigkeit, Generosität, Grossmut, Grosszügigkeit, Hochherzigkeit, Konzilianz. Back-translating this set of German terms generates the following set of English terms: bounteousness, bounty, broad-mindedness, catholicity, conciliatoriness, courtesy, freedom, gallantry, generosity, generousness, high-mindedness, liberality, liberalness, lordliness, magnanimity, magnanimousness, mobility, noble mindedness, nobleness, permissiveness, promiscuity. Therefore, when creating pairs of English–German value terms, sometimes also less common terms in either language entered our set, as a common term was already represented as a translation of yet another term.

were considered to be included in the list that were indeed value terms, that were specific enough, and for which a new counterpart in the other language was available. They were included either as addendum (18 in total) or as a replacement of a less common term.[3] In total, we had 12 replacements for English and 7 replacements for German terms. Second, we checked for terms whose word bags contained less than 5 terms or whose word bags contained terms that were not present in any other word bag (suggesting that they would have no similarity with any other term, see next section) in both languages. This was the case for 6 terms, and they were thus excluded from the list. By this refinement we obtained a set of value terms that contained 460 words per language (the list of all terms and their translation is provided as Supplementary information). This number that resulted from our search strategy is certainly contingent, but it would be no problem to add further value terms in the methodology described further below.

Thus, for each of the 460 terms a semantic constellation $(t, \{t, t_{syn1}, t_{syn2}, t_{syn3}, \ldots, t_{syn\,n}\})$ was available. All word bags together consisted of a total of 3749 distinct English terms and 4775 distinct German terms. That is, a German value term was, in the mean, associated with more synonyms.

2.1.2. Step 2 – Iterative Data Analysis Procedure

The goal of the data analysis was to identify groups of value terms that can be considered to be sufficiently similar for forming a "value cluster", i.e. a set of value terms that are similar enough assuming a pre-theoretical understanding of the terms. In this way we wanted to reduce the large number of terms such that the resulting map is easier to read and use. The procedure had two iterations: the first one served to identify value clusters and the second one served to refine the clusters found. A single iteration consisted in machine recognition and expert evaluation. There are several reasons why this method cannot solely rely on a machine classification approach: First, Thesaurus databases are fuzzy by nature. There is no standard procedure why certain terms qualify as synonyms for another term. This fuzziness likely leads to classification errors that require human intervention. Second, the number of elements in a word bag per term varies greatly. In the machine recognition step this can result in an obvious misplacing of a term because of the small size of its word bag. Here, semantic expertise is required for finding the "correct neighbors" of those terms in order to allow for a plausible reduction of the number of terms. Third, there is no "natural" cut-off when the reduction of terms can be considered complete. For this, a human expert decision is required. For supporting the human experts, we use an important heuristic in the method that consisted of a visualization tool. This tool (Superparamagnetic Agent Maps, SAM) is presented in Section 3. We now describe the two iterations in some more detail:

- *Iteration 1 – machine classification*: The first iteration step started with machine classification using SAM. Here, the aim was to optimize the map such that it best represents the real similarities in the high-dimensional value term space. After each mapping, a refinement procedure was implemented as follows: We calculated for each value term X, which values were the closest 10 neighbors (in terms of the Euclidean distance) within the 2-dimensional representation generated by the map and we compared this set with the closest 10 neighbors according to the dissimilarity matrix. For all values of the intersection of these two sets, the dissimilarity between those values and X was halved, i.e. the "attractor" effect inherent in the similarity measure is enhanced. This procedure generated a refined dissimilarity matrix such that in the 2-dimensional representation, values that were close to each other in the original high-dimensional space moved closer to each other on the map. This iterative procedure was repeated for each language until the distribution of all dissimilarities in iteration step $s+1$ was not any more significantly different from the distribution in step s (Mann–Whitney-Test). For the English and German value map, this was the case for $s = 3$ and $s = 4$, respectively. Finally, using the sequential superparamagnetic clustering algorithm (SSC; Ott, Kern, Steeb, & Stoop, 2005), we identified all value clusters for each iteration step in the original data space. This allowed identifying three types of value clusters: (1) Stable value clusters containing values that grouped together across all iteration steps; (2) unstable value clusters identified using the dissimilarity matrix of the last iteration step; (3) likely clusters of values that have been identified solely based on the 2-dimensional representation, i.e., the values were close to each other on the value map, but have not been identified as a cluster using SSC. The output was a value map containing all 460 values, in which the groups identified have been revolved. A few values that were identified by SSC to belong to a specific group but that were far away on the map were moved manually closer to the group on the 2-dimensional representation (recall that one cannot expect that a 2-dimensional representation of a very high-dimensional space is perfect, i.e. it does not display all values close to each other that are indeed close in the original space). Overall, both the English and German value maps contained 41 value groups that included 323 specific value terms in each language.

[3] An example of a replacement is the term "diligence" (German translation: "Fleiss", "Sorgfalt" or "Gewissenhaftigkeit") that was on the list but did not appear in the synonym word bags, whereas the term "carefulness" was 18 times proposed as synonym of other terms in the list. Therefore, we replaced the term "diligence" by "carefulness" as English translation of the German "Sorgfalt". We remind the ambiguity of translations: "diligence" can also be translated as "Fleiss" – but this German term already had another English counterpart ("industry") that was more common in the synonym set.

– *Iteration 1 – expert evaluation:* The English and German value maps identified in this way were then evaluated by six experts per language. The experts had a background in Philosophy, Psychology or English/German language studies. They were informed on how the map was generated and they were instructed to do the following evaluation for each group: First, verify whether a suggested group contained value terms that should not belong to the group; second, test whether a suggested group should be sub-divided into smaller groups; third, examine whether there are other values that either have not been attributed to any group, or that are contained in another group; fourth, choose a value that provides the best name for the group. To provide assistance, the value map served as heuristic since values that are close to each other usually share some similarity, i.e. the experts were not forced to consult a list of 460 values but they could work with a visual representation allowing for an efficient check of the machine suggestion.
– *Iteration 2 – machine classification:* Expert feedback was used to refine the map as follows: Groups were confirmed when five of six experts agreed that the values forming the group indeed "belonged together" (i.e. demonstrated a similar meaning). We also validated the results across languages (i.e. we identified groups such that the translations of their representatives were also identified as a group in the other languages). In this way, 78 across-language groups were identified that contained 247 value terms. Then, a second machine classification step was designed to attribute the remaining value terms to the groups identified as follows: For each value group, the synonyms of each value forming the group were merged and we calculated for all remaining 213 value terms (for each language) to which group they fit best by using our similarity measure. Each value then was attributed to the group to which it fit best. In this way, we generated for each language a list that showed the 78 core groups, the value terms of which the majority of the experts believed that they were the best names of the group, and the residual values that have been attributed to each group.
– *Iteration 2 – expert evaluation*: The result of the second machine classification step was then again presented to experts. Those had the same profile as in the first iteration: Six experts per language provided an evaluation, whereas three out of six came from the expert pool of the first iteration and three other experts newly joined the study. After outlining the procedure used in the second machine classification step, the experts were asked to evaluate the lists provided to them as follows: They were asked to examine whether they agreed in the allocation of the residual values or whether they would prefer relocating some of them. Furthermore, they should generally evaluate the suggested grouping, e.g., whether they think that some groups are superficial, incorrectly named or whether there are some basic values missing.

2.1.3. Step 3 – Value Map Finalization

The second expert feedback served to finalize the grouping. For each value group, only those residual values for which all six experts agreed upon have been kept in the group. When there was disagreement (74 values in English, 38 values in German), we calculated for each group the distribution of all in-group similarities and compared this with the distribution of the similarities between the group-values and the test value. Only when the distributions were not statistically significant distinct (Mann–Whitney-Test, $p > 0.05$) was the test value kept in the group. To identify particularly weak groups, we calculated for each group the distribution of all inter-group similarities and compared it with the similarity distribution of all between-group values. This revealed that the "competition" (German: "Wettbewerb") cluster was very weak, so we decided to disintegrate this group and we kept "competition" ("Wettbewerb") as a one-value group. This decision was made due to theoretical reasons, as competition is an often discussed value orientation in economic discourses and should therefore be represented in the value map. In this way, 32 English values and 14 German values remained unattributed. Finally, we merged the synonyms of all values that formed a specific group and we calculated the overlap of the synonym sets of the remaining values with the synonym set of each group. If there was a clear indication that a remaining value fit strongly to one group,[4] the value was added to that group. At the end, we had identified 78 groups composed of in total 440 (English) and 449 (German) terms; 20 English and 11 German values remained unattributed.

Based on this result, we created the SAM for the 78 groups, whereas the similarity relation has been calculated using the merged word bags of the values that form a single group. This has been done for each language. Since such a map will never precisely display the real topology of the original, high-dimensional space, we calculated for each point on the map the sum of the differences between the point and all its neighbors both in the map and in the original space (normalized to the longest dissimilarity/distance in either case; map aberration index, see Ott et al., 2014 for further details). The smaller this sum, the better the map displays the real distance distribution of a point from its neighbors in the original space, so this number is a proxy for the quality of the map. To increase the heuristic value of the maps, we rescaled the sizes of the points themselves so that larger points indicate greater topological certainty.

2.2. Survey studies for testing the value group composition

Our approach is based on the assumption that the entries in thesaurus databases reflect similarities of terms

[4] We identified the group that had the highest similarity to a residual value. If the similarity value is at least twice as big as the mean pairwise similarity of all values that already belong to the group and if there was no explicit expert statement that opposed attribution, the residual value was included into the group.

based on their usage in (written) language in a rather broad language community, neglecting differences e.g., due to different dialects. But it is not clear whether the result of our analysis that is based on this assumption leads to value groups that people indeed consider to be "similar" assuming a pre-theoretical understanding of the terms. Take the example of "empathy" and "sympathy" that are in the same group, but are rather different concepts from an elaborated theoretical point of view (Eisenberg, 2000). However, our approach does not aim to be this fine-grained. Instead, we wanted to identify "sufficiently" similar relations among terms that reflect the use of these terms in natural language.

We performed two online survey studies – one in English (USA) and one in German (German part of Switzerland) – designed to empirically test the extent to which participants consider the value terms of a single group to be synonyms of the same concept of a value. For this purpose, the participants had to evaluate terms that were presented as being synonyms of a specific value concept. They were asked to decide which terms belong to the same value category, and then (by using the mouse pointer) to drag all words that do not fit to their understanding of the value to a "trash bin". The remaining words then should represent the value based on their choices. The study was cleared in accordance with the ethical review processes of the University of Zurich and within the "Ethical Guidelines for Psychologists of the Swiss Society for Psychology".

For example, the participants were presented with the following list of terms (in an alphabetical order): aggressiveness, altruism, benevolence, charity, kindheartedness, kindness, philanthropy, sacrifice, selflessness, and thoughtfulness. By using their mouse, they were asked to drag all words that did not fit into the box "trash", such that the remaining words form, according to their opinion, a synonym set circumscribing a value. One term ("aggressiveness" in this case) was a distractor term, i.e., a term that was surely not a synonym to the other group members and that was expected to be dragged to the trash bin. Participants that did not exclude distractor terms were considered uncooperative participants and were excluded from the analysis.

As the sizes of the groups identified in the thesaurus analysis differ largely, we added to the smaller groups (those that had less than 6 terms) additional synonyms from the synonym bag of the group for adjusting the group sizes. In total (including the distractor terms), the groups used in the surveys consisted of 7–14 terms. Participants rated a random sub-sample of groups (12 in the English, 10 in the German survey).

Overall, in the English and German survey, we included data of 280 English and 201 German participants, respectively. The survey among participants for the English value groups was conducted from September 26 to September 29, 2013 using *Amazon Mechanical Turk* (restricted to US Americans). In this sampling method, participants received a small reimbursement for completion of the survey. The survey for the German value groups was conducted in February 2014 using a service provided by the University of Zurich that addresses students, faculty and staff of the university (i.e., Swiss-German speaking). Here, a lottery (*Amazon* gift cards) was chosen as incentive for participation.

3. Theory and calculation

3.1. Similarity measure

An important heuristic in the method outlined above consisted of a visualization tool to help the experts in evaluating the clusters and to create the final maps. The idea was to map the similarities between the values terms represented by word bags on a two dimensional plane such that similarities translate into distances on the map. Therefore, a similarity measure is needed as input for the visualization tool.

For comparing semantic constellations of the value terms, we created a similarity measure by calculating the relative overlap of each pair of word bags. Let $T = \{t, t_{syn1}, t_{syn2}, t_{syn3}, \ldots, t_{syn\,n}\}$ be the word bag of a term t. Then, the similarity S of two terms t^1 and t^2 is defined as:

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

The similarity of two terms equals one, when the word bag of one term overlaps completely with the word bag of another term. The similarity is 0, if the word bags are mutually distinct. In this construction, similarities between two terms t_1 and t_2 are *a priori* more likely if $T^1 \ll T^2$ (or vice versa), i.e. a term with few synonyms has a higher probability to be similar to a term with many synonyms versus a situation where the word bags have comparable size. This means that terms with large word bags tend to act as "attractors" in the iterative procedure described below. Although S is a similarity and not a distance measure,[5] it is sufficient for our visualization method (see Ott et al., 2014 for details). This allowed for generating a dissimilarity matrix (1 minus the similarity value) that served as input for the superparamagnetic agent map (SAM) algorithm (see below).

To empirically validate the similarity maps obtained from thesaurus data, we needed an analogous measure of similarity for participant-generated data. Here, we calculated the ratio of how often two terms were classed together in the same group by the participants compared to the total number of participants who were confronted

[5] A distance measure has to fulfill the triangle inequality, i.e. $d(a,b) + d(b,c) \geqslant d(a,c)$, where $d(a,b)$ stands for the distance between points a and b. This inequality is violated by our measure $S(t^1, t^2)$. In the following, we use the term "dissimilarity" to denote $1 - S(t^1, t^2)$ in the original space, whereas on the map, we use the term "distance", as we refer to the Euclidean distance of points on the two-dimensional plane.

with such a comparison. Although the similarity measures obtained in this way are not expected to match the ones of thesaurus similarity (e.g. it is much less likely to obtain a similarity value of 0 in the survey), we expect that the measures correlate to the extent that the participant ratings reflect thesaurus similarity.

3.2. Superparamagnetic Agent Map

Visualizing a semantic space is a standard problem of dimensionality reduction. Classical approaches aim to represent the data structure on a linear subspace of the original data space. For example, principal component analysis performs a projection onto the axes with maximal data variance; whereas the goal of multidimensional scaling is to find a low-dimensional embedding that preserves the inter-point distances. These methods often perform poorly when applied to nonlinear data structures. Furthermore, for many real-world applications, data vectors are not available. Instead, researchers are faced with similarity or proximity data, as in our case. We therefore have applied a novel visualization tool called Superparamagnetic Agent Mapping (SAM) (Ott et al., 2014).

To conceptualize this mapping, imagine each term as a particle that naturally repels all other particles. However, as overlap between two terms increases, they become more attracted to each other. Thus, SAM typically produces clustering, where several particles clump together (connoting similarity) while collectively repelling a different cluster (connoting collective difference between the two clusters). The notion of "superparamagnetic" refers to a form of magnetism, where the overall magnetization is zero, but small ferromagnetic clusters occur in which magnetization can randomly flip direction under the influence of temperature. This model has been transformed into the so-called "superparamagnetic clustering algorithm" (Blatt, Wiseman, & Domany, 1996) that does not assume any structure of the underlying distribution of the data. This algorithm has been further advanced by one of the authors (Ott et al., 2005) and provides the theoretical basis of our visualization tool.

More formally, the method is an iterative two-step procedure that is repeated until a threshold condition is reached. In the first step, each data item (=synonym or antonym term) is represented by a Potts spin variable and the dissimilarity matrix is encoded in the spin couplings. The spin system is treated in the formalism of the canonical ensemble, giving the probability for a certain spin configuration. One then can observe that the spins whose corresponding data items are similar tend to cluster in terms of the pair correlation G_{ij}, i.e., the probability of two spins being in the same state. 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

In the second step, each data item is represented by an agent in a 2-dimensional coordinate system. The agents move according to laws that are governed by the local interactions of the spin system calculated in step 1. In order to calculate G_{ij}, a Markov chain Monte Carlo algorithm needs to be employed, which generates a sequence of binary pair correlation states $G_{ij}(t) \in \{0, 1\}$. Starting from a random distribution, two agents move towards each other if $G_{ij} = 1$, i.e., if the corresponding spins are in the same state in the current configuration, otherwise the agents drift apart, leading to a 2-dimensional distribution of agents. For the precise formalism, we refer to Ott et al. (2014). It has been shown (Ott et al., 2014) that SAM is superior to standard methods such as factor analysis, principal components analysis, and multidimensional scaling in preserving the topology of the data space with clustered data.

4. Results

4.1. Thesaurus analysis

As to the results of the thesaurus analysis, the composition of the value groups is outlined in Table 1; the value maps for the 78 value groups are displayed in Figs. 2 and 3.

Table 1 reveals that the group sizes differ substantially: the largest group (entitled: "joy") contains 13 terms, whereas other groups have only few terms ("competition" is a special case, as outlined in Section 2). The terms printed in bold are the group names as suggested by the experts. Recall that each term stands for a word bag that includes 5–60 terms in English (mean: 25, median: 23 terms) and 5–168 terms in German (mean: 46, median: 35 terms). As expected, there are also language-specific differences in the composition of the value groups (columns "Language-typical elements English/German"). This results from the fact that the word-bags of each term in either language differ and that there is some ambiguity in translating terms from English to German.

The Table also outlines how often terms were attributed to the groups in the survey study: the number in brackets shows the relative frequency a term was kept as a member of the group (see Section 4.2. for further information).

It is important to keep in mind that the groups contain terms that are similar only to some degree in terms of the used similarity measure. The mean overlap of word bags in single groups (intra-group) varied quite substantially between 3.5% and 91.7% (the mean overlap over all groups is 29.8%) in English and between 7.5% and 100% in German (mean over all groups: 40.6%). Nevertheless, the inter-group similarities of single values (i.e., the overlap of word bags of a value X of group A and a value Y of group B) are much smaller (mean: 1.6%), so that our algorithm as described in Section 2 still is able to discern groups that have a rather low thesaurus similarity but are accepted as groups by the experts.

Table 1
Value group composition: The terms printed in bold are the group names used for the value maps. The number in brackets indicate the relative frequency by which a term as been kept in the group in the survey study (1: the term was never deleted; 0: the term was always deleted).

Language-typical elements English	English values with German translations	German values with English translations	Language-typical elements German
	aggressiveness (0.93), pugnaciousness (0.89), toughness (0.87)	**Aggressivität (0.74)**, Härte (0.65), Kampfeslust (0.87)	Herausforderung (0.65), Widerstandsfähigkeit (0.48)
benevolence (0.91), kindheartedness (0.98), kindness (0.98), thoughtfulness (0.98)	**altruism (0.92)**, charity (0.96), philanthropy (0.87), sacrifice (0.87), selflessness (0.96)	**Altruismus (0.84)**, Aufopferung (0.60), Nächstenliebe (0.92), Philanthropie (0.76), Selbstlosigkeit (0.76)	Gastfreundschaft (0.60), Hilfsbereitschaft (0.96)
realism (0.76)	**authenticity (0.95)**, clarity (0.70), truth (0.92), truthfulness (0.97	**Authentizität (0.90)**, Klarheit (0.87), Wahrhaftigkeit (0.93), Wahrheit (0.87)	Gewissheit (0.61), Glaubwürdigkeit (0.97), Prägnanz (0.48)
sovereignty (0.66)	**authority (0.96)**, influence (0.96), persuasion (0.94), power (0.96)	**Autorität (0.88)**, Einfluss (0.96), Macht (0.56), Überzeugungskraft (0.76)	Bestimmtheit (0.88), Durchsetzungsvermögen (0.88), Kaltschnäuzigkeit (0.20), Signifikanz (0.48)
	autarchy (0.52), **autonomy (0.91)**, independence (0.98), self-determination (0.95), self-reliance (0.98)	Autarkie (0.67), **Autonomie (0.92)**, Eigenständigkeit (0.96), Selbständigkeit (1.0), Selbstbestimmung (0.96)	Bewegungsfreiheit (0.63), Eigenverantwortung (0.96), Mündigkeit (0.75), Souveränität (0.75)
attraction (0.75), grace (0.93)	**beauty (0.91)**, charm (0.98), elegance (0.93), poise (0.80)	Charme (0.83), Eleganz (0.89), Haltung (0.78), **Schönheit (0.78)**	Glanz (0.78), Raffinesse (0.39)
connection (0.52), presence (0.34)	**care (1.0)**, concern (0.93), nurture (0.95), support (1.0)	**Fürsorge (0.97)**, Pflege (0.87), Unterstützung (1.0), Zuwendung (0.73)	
conscientiousness (0.98), deliberateness (0.85), focus (0.72), sensitivity (0.21)	**carefulness (0.98)**, meticulousness (0.94), scrupulousness (0.70), thoroughness (0.96)	Akribie (0.77), Gewissenhaftigkeit (0.92), Gründlichkeit (0.92), **Sorgfalt (1.0)**	
	chastity (0.89), innocence (0.92), virginity (0.84)	Jungfräulichkeit (0.81), **Keuschheit (0.81)**, Unschuld (0.95)	
	civility (1.0), considerateness (0.95), propriety (0.46), tact (0.76)	Anstand (0.79), Fingerspitzengefühl (0.79), **Höflichkeit (0.71)**, Rücksichtnahme (1.0)	Vorsicht (0.54)
effort (0.96), keenness (0.69)	**commitment (0.88)**, drive (0.96), enterprise (0.65), industry (0.41)	**Engagement (1.0)**, Fleiss (0.94), Tatkraft (0.89), Unternehmungsgeist (0.72)	Betriebsamkeit (0.72), Einsatz (0.94)
	clemency (0.78), **compassion (0.98)**, mercy (1.0), pity (0.83)	Barmherzigkeit (0.87), Milde (0.94), Mitgefühl (1.0), **Mitleid (0.81)**	Warmherzigkeit (0.94)
	competition (1.0)	**Wettbewerb (1.0)**	
comfort (0.98)	**contentment (0.93)**, coziness (0.90), relaxation (0.93), satisfaction (0.85)	**Behagen (0.80)**, Entspannung (0.75), Gemütlichkeit (0.65), Zufriedenheit (1.0)	Lebensqualität (0.90), Leichtigkeit (0.80), Wohlbefinden (1.0)
helpfulness (0.96), reciprocity (0.73), synergy (0.67)	assistance (0.98), **cooperation (0.98)**, help (0.93)	Beistand (0.78), Hilfe (0.91), **Kooperation (0.69)**	
confidence (0.91)	bravery (1.0), **courage (1.0)**, fearlessness (0.98), guts (0.96), intrepidity (0.77), nerve (0.85), pluck (0.49)	Courage (0.95), Furchtlosigkeit (0.90), Mumm (0.75), Schneid (0.80), **Tapferkeit (0.95)**, Unerschrockenheit (0.95)	Entscheidungskraft (0.60), Kraft (0.55), Mut (1.0), Willenskraft (0.70)
cleverness (0.90)	**creativity (0.96)**, imagination (0.98), ingenuity (0.94), inspiration (0.96), inventiveness (1.0), originality (0.96), resourcefulness (0.84)	Einfallsreichtum (1), Erfindungsgabe (0.88), Findigkeit (0.80), Inspiration (0.84), **Kreativität (1.0)**, Originalität (1.0), Phantasie (0.96)	Ausdrucksfähigkeit (0.64), Extravaganz (0.24), Intuition (0.68), Produktivität (0.28)
decency (0.90), merit (0.80), value (0.82), worth (0.82)	**dignity (0.90)**, majesty (0.33), nobility (0.61), self-respect (0.80)	Edelmut (0.58), Majestät (0.37), Selbstachtung (0.74), **Würde (1.0)**	Selbstvertrauen (0.42)

dynamism (0.52)	control (0.94), **dominance (0.92)**, leadership (0.96), superiority (0.81)	**Dominanz (0.72)**, Führung (0.91), Kontrolle (0.86), Überlegenheit (0.64)	
productivity (0.92)	effectiveness (0.97), efficacy (0.79), **efficiency (1.0)**	Effektivität (0.84), **Effizienz (0.95)**, Wirksamkeit (0.68)	Errungenschaft (0.26)
understanding (0.96)	**empathy (0.92)**, sympathy (0.92)	**Empathie (0.95)**, Sympathie (0.55)	Sensibilität (0.82), Teilnahme (0.72), Zuneigung (0.41)
eagerness (0.90), zeal (0.92)	ardor (0.72), **enthusiasm (0.95)**, fervor (1.0), intensity (0.90)	Begeisterung (1.0), Eifer (1.0), **Enthusiasmus (1.0)**, Intensität (0.72)	Anstrengung (0.44), Elan (1.0), Fitness (0.28), Lebhaftigkeit (0.61), Temperament (0.61)
evenness (0.97), fair-mindedness (0.92), neutrality (0.84), objectivity (0.68)	**equality (0.97)**, equity (0.79)	Billigkeit (0.08), **Gleichheit (0.92)**	Konformität (0.71), Kongruenz (0.50)
performance (0.50)	brilliance (0.93), completion (0.13), **excellence (0.98)**, exceptionality (0.93), perfection (0.89)	Aussergewöhnlichkeit (0.78), Brillanz (0.89), Erstklassigkeit (0.85), **Perfektion (0.78)**, Vollendung (0.78)	Einzigartigkeit (0.85), Maximum (0.59), Meisterschaft (0.56)
	evenhandedness (1.0), **fairness (1.0)**, impartiality (0.82), justice (0.82)	Fairness (1.0), **Gerechtigkeit (1.0)**, Unparteilichkeit (0.90), Unvoreingenommenheit (0.85)	Angemessenheit (0.65), Anständigkeit (0.80), Rechtschaffenheit (0.80), Unbestechlichkeit (0.80)
certainty (0.95)	conviction (0.87), **faith (0.67)**, trust (0.74)	**Glaube (0.95)**, Überzeugung (0.95), Vertrauen (0.89)	
amenability (0.71), mobility (0.65)	adaptability (1.0), **flexibility (1.0)**	Anpassungsfähigkeit (0.96), **Flexibilität (1.0)**	
amity (0.72), camaraderie (0.94), hospitality (0.94), sociability (0.96), warmth (1.0)	affability (0.77), bonhomie (0.60), **friendliness (1.0)**, geniality (0.87), gentleness (0.68)	**Freundlichkeit (1.0)**, Gutmütigkeit (0.96), Herzlichkeit (0.96), Liebenswürdigkeit (0.84), Umgänglichkeit (0.68)	Wohlwollen (0.80)
liberality (0.42)	forgiveness (0.93), **generosity (0.72)**, lenience (0.88), tolerance (0.88)	**Grosszügigkeit (0.50)**, Nachsicht (0.85), Toleranz (0.95), Vergebung (0.70)	Erkenntlichkeit (0.30), Geduld (0.70), Gnade (0.65), Güte (0.90), Herzensgüte (0.85), Kulanz (0.65), Sanftmut (0.90), Verständnis (1.0)
congruence (0.71), universality (0.61)	**harmony (0.97)**, nonviolence (0.63), peace (0.87), unity (1.0)	Eintracht (0.90), Frieden (0.95), Gewaltlosigkeit (0.55), **Harmonie (0.90)**	
	audacity (0.84), boldness (0.95), **heroism (0.82)**, temerity (0.55)	**Heldenmut (0.67)**, Kühnheit (0.83), Verwegenheit (0.67), Wagemut (0.94)	
incorruptibility (0.58)	directness (0.95), **honesty (1.0)**, straightforwardness (0.95), transparency (0.95)	Direktheit (0.76), **Ehrlichkeit (0.95)**, Geradlinigkeit (0.62), Transparenz (0.81)	Aufgeschlossenheit (0.33), Aufrichtigkeit (0.95), Geradheit (0.81)
conscience (0.87), good (0.83), goodness (0.87), morality (0.94), repute (0.43), righteousness (0.79), uprightness (0.85)	**integrity (0.98)**, probity (0.43), sincerity (0.81), virtue (0.94), virtuousness (0.91)	**Integrität (0.90)**, Lauterkeit (0.86), Redlichkeit (1.0), Tugend (0.57), Tugendhaftigkeit (0.87)	Ernsthaftigkeit (0.57)
intuition	acumen (0.82), comprehension (0.95), discernment (0.85), insight (1.0), **intelligence (1.0)**, rationality (0.90), reason (0.95), sagacity (0.67)	Auffassungsgabe (0.76), Einsicht (0.81), **Intelligenz (0.86)**, Rationalität (0.86), Scharfsinn (0.81), Urteilsfähigkeit (0.95), Urteilsvermögen (1.0), Vernunft (0.86)	Bewusstheit (0.81)
	bliss (0.94), cheerfulness (0.92), exhilaration (0.88), fun (0.94), gladness (0.94), happiness (1.0), hedonism (0.39), hilarity (0.73), humor (0.80), jocularity (0.78), **joy (0.98)**, playfulness (0.84), pleasure (0.96)	**Freude (0.90)**, Fröhlichkeit (1.0), Frohsinn (0.95), Genuss (0.62), Glück (0.86), Glückseligkeit (0.95), Heiterkeit (1.0), Hochgefühl (0.90), Humor (0.71), Munterkeit (0.90), Spass (0.86), Vergnügen (0.90), Witzigkeit (0.62)	Geselligkeit (0.71)

(*continued on next page*)

Table 1 (*continued*)

Language-typical elements English	English values with German translations	German values with English translations	Language-typical elements German
	compliance (0.85), law (0.94), **lawfulness (1.0)**, legality (0.98), right (0.73)	Gesetz (0.91), Legalität (0.96), Recht (0.82), **Rechtmässigkeit (0.91)**, Regelkonformität (0.96)	
	freedom (0.97), **liberty (1.0)**	**Freiheit (0.96)**, Unabhängigkeit (0.96)	
affection (0.91), dedication (0.91)	devotion (1.0), **love (0.82)**	Hingabe (0.90), **Liebe (0.95)**	Innigkeit (0.80)
attachment (0.82), patriotism (0.56), steadfastness (0.87)	adherence (0.82), faithfulness (1.0), fidelity (0.90), **loyalty (0.97)**	Anhänglichkeit (0.38), Ergebenheit (0.73), **Loyalität (0.96)**, Treue (0.88)	Dienst (0.65), Gefolgschaft (0.85)
introversion (0.49), reflection (0.74), regard (0.77)	concentration (0.64), consideration (0.85), **mindfulness (0.97)**, regardfulness (0.79)	**Achtsamkeit (0.90)**, Aufmerksamkeit (1.0), Konzentration (0.70), Rücksicht (0.85)	Behutsamkeit (0.75), Fokus (0.50), Wachsamkeit (0.80)
	frugality (0.58), humility (1.0), **modesty (1.0)**, simplicity (0.69)	Anspruchslosigkeit (0.75), **Bescheidenheit (1.0)**, Einfachheit (0.95), Genügsamkeit (1.0)	Natürlichkeit (0.30)
	ambition (1.0), aspiration (1.0), impetus (0.66), **motivation (1.0)**	Antrieb (1.0), Ehrgeiz (0.87), **Motivation (1.0)**, Strebsamkeit (0.87)	Interesse (0.87), Wissensdurst (0.83)
	nonmaleficence (0.26), treatment (0.69)	Behandlung (0.89), **Nichtschaden (0.47)**	Versorgung (0.79)
conformity (0.76), deference (0.81), meekness (0.81)	humbleness (0.83), **obedience (0.86)**, willingness (0.74)	Bereitwilligkeit (0.95), Demut (0.25), **Gehorsam (0.30)**	Abhängigkeit (0.10), Beflissenheit (0.70), Bereitschaft (1.0), Duldsamkeit (0.50), Einsatzbereitschaft (0.95)
	dispassion (0.78), neutrality (0.93), **objectivity (0.68)**	Neutralität (0.73), **Objektivität (0.95)**, Sachlichkeit (0.95)	Realismus (0.58)
inquisitiveness (1.0), interest (0.94), openmindedness (0.91)	curiosity (1.0), informality (0.26), **openness (0.85)**, spontaneity (0.53)	Neugier (0.78), **Offenheit (1.0)**, Spontanität (0.83), Ungezwungenheit (0.83)	Freizügigkeit (0.26), Vielfalt (0.39), Zugänglichkeit (0.82)
idealism (0.72)	hope (0.93), **optimism (0.98)**	**Hoffnung (0.94)**, Optimismus (0.94)	Frohmut (0.71), Trost (0.12)
intimacy (0.88)	delight (0.83), ecstasy (0.95), lust (0.71), **passion (0.93)**, sensuality (0.95), sexuality (0.83), stimulation (0.80)	Aufregung (0.84), Ekstase (0.89), **Leidenschaft (1.0)**, Lust (1.0), Sexualität (0.79), Sinnlichkeit (o.79), Wonne (0.79)	
allegiance (0.73)	**piety (0.77)**, religiousness (0.75)	**Frömmigkeit (0.95)**, Religiosität (0.91)	
correctness (0.96)	accuracy (1.0), exactness (0.96), **precision (0.98)**, rigor (0.12)	Exaktheit (0.92), Genauigkeit (0.96), **Präzision (0.92)**, Strenge (0.36)	
	confidentiality (1.0), **privacy (0.98)**, secrecy (0.90)	Diskretion (0.90), **Privatsphäre (1.0)**, Verschwiegenheit (0.75)	
preparedness (0.84), qualification (0.86)	competence (1.0), expertise (0.98), **professionalism (0.70)**	Expertise (0.95), Kompetenz (1.0), **Professionalität (0.80)**	
acquisition (0.82), ownership (0.80)	gain (0.96), **profit (0.96)**, profitability (0.94), profitableness (0.92), yield (0.80)	Einträglichkeit (0.93), Gewinn (1.0), **Profitabilität (1.0)**, Rendite (1.0), Rentabilität (1.0)	Kommerz (0.90), Verdienst (0.83), Vorteil (0.72)
maturation (0.88)	growth (1.0), **progress (0.98)**	**Fortschritt (0.86)**, Wachstum (0.95)	
achievement (0.64), benefit (0.62), extravagance (0.84), wellbeing (0.56)	convenience (0.29), fortune (0.84), luxury (0.84), **prosperity (0.96)**, richness (0.89), wealth (0.89)	Komfort (0.90), Luxus (0.90), Reichhaltigkeit (0.57), Reichtum (1.0), Vermögen (0.95), **Wohlstand (0.86)**	Gut (0.48)
calculation (0.58), caution (0.95), cautiousness (0.98)	discretion (0.95), foresight (0.78), **prudence (0.88)**	**Besonnenheit (0.95)**, Umsicht (0.95), Weitsicht (0.80)	Abgeklärtheit (0.40), Bedächtigkeit (0.65), Bedachtsamkeit (0.95), Beschaulichkeit (0.40)
	cleanliness (1.0), hygiene (1.0), **purity (0.79)**	Hygiene (1.0), **Reinheit (0.71)**, Reinlichkeit (1.0)	

assertiveness (0.79), continuity (0.27), decisiveness (0.85), determination (1.0), endurance (0.85), willpower (0.85)	perseverance (0.98), **purposefulness (0.85)**, resoluteness (0.94), tenacity (0.92)	Beharrlichkeit (1.0), Entschiedenheit (0.91), **Entschlossenheit (0.95)**, Hartnäckigkeit (0.95)	Standhaftigkeit (0.91), Unerschütterlichkeit (0.86)
acceptability (0.35), fame (0.83)	image (0.85), popularity (0.95), prestige (0.90), prestigiousness (0.85), renown (0.95), **reputation (0.95)**, status (0.93)	Ansehen (1.0), Beliebtheit (0.76), Image (0.95), Prestige (0.90), Rang (0.81), **Reputation (1.0)**, Ruf (1.0)	Format (0.24), Geltung (0.86), Leumund (0.57), Renommee (0.95)
gratitude (0.85), thankfulness (0.76)	admiration (0.92), affirmation (0.68), appreciation (0.95), esteem (0.76), recognition (0.98), **respect (0.98)**, reverence (0.83), worship (0.56)	Anerkennung (0.95), Bestätigung (0.47), Bewunderung (0.37), Ehrfurcht (0.32), Hochachtung (0.58), **Respekt (0.95)**, Verehrung (0.21), Wertschätzung (1.0)	Achtung (0.84), Akzeptanz (0.68), Beachtung (0.79)
kudos (0.71)	honor (1.0), **respectability (0.94)**	**Ehre (0.65)**, Ehrenhaftigkeit (0.20)	Erfolg (0.95), Ruhm (0.90)
	accountability (0.95), dutifulness (0.80), duty (0.80), **responsibility (0.95)**	Pflicht (0.81), Pflichttreue (0.81), Verantwortlichkeit (0.76), **Verantwortung (0.87)**	Gewissen (0.90), Moral (0.71), Pflichteifer (0.43)
earnestness (0.38)	holiness (0.93), **sanctity (0.95)**, sublimity (0.68)	Erhabenheit (0.94), **Heiligkeit (0.71)**, Unverletzlichkeit (0.71)	Pietät (0.53)
resistance (0.24), vigilance (0.78)	protection (0.95), safety (0.92), **security (0.97)**	Geborgenheit (1.0), Schutz (1.0), **Sicherheit (1.0)**	
	discipline (0.96), restraint (0.98), **self-control (0.96)**, self-discipline (1.0)	Beherrschung (0.68), Disziplin (0.95), **Selbstbeherrschung (0.95)**, Selbstdisziplin (1.0)	
coolness (0.91), ease (0.85), imperturbability (0.70), mellowness (0.91), moderation (0.57), patience (0.91), tranquility (0.98)	calmness (1.0), equanimity (0.64), **serenity (0.98)**	**Gelassenheit (1.0)**, Gleichmut (0.43), Ruhe (0.86)	Ausgeglichenheit (1.0)
adequacy (0.49), articulateness (0.76), finesse (0.84), hustle (0.27), mastery (0.97), readiness (0.65), speechcraft (0.54)	adroitness (0.62), proficiency (0.95), quickness (0.46), **skill (0.97)**	Flinkheit (0.83), **Geschicklichkeit (0.96)**, Gewandtheit (0.96), Können (0.71)	Qualifikation (0.42)
	affiliation (0.86), community (0.93), family (0.88), partnership (0.93), **solidarity (0.74)**	Familie (0.95), Gemeinschaft (1.0), Partnerschaftlichkeit (0.95), **Solidarität (1.0)**, Zugehörigkeit (0.95)	Bindung (0.91), Freundschaft (1.0), Gegenseitigkeit (0.77), Kameradschaft (0.91), Synergie (0.64), Verbindung (0.77)
significance (0.91)	meaning (0.93), **spirituality (0.52)**, transcendence (0.54)	Lebenssinn (0.79), **Spiritualität (0.79)**, Transzendenz (0.76)	Introversion (0.62), Reflektion (0.97), Tiefe (0.97)
consistency (1.0)	constancy (0.92), mettle (0.18), order (0.84), permanence (0.92), **stability (0.97)**, structure (0.87), tradition (0.84)	Beständigkeit (0.95), Konstanz (0.79), Ordnung (0.32), **Stabilität (1.0)**, Standfestigkeit (0.74), Struktur (0.47), Tradition (0.53)	Kontinuität (0.84)
fitness (0.71), force (0.88), vigor (0.95)	capacity (0.40), potency (0.83), **strength (1.0)**	**Leistungsfähigkeit (1.0)**, Potenz (0.64), Stärke (0.73)	Ausdauer (0.86), Leistung (0.82)
	preservation (1.0), **sustainability (0.93)**	**Nachhaltigkeit (0.91)**, Umweltschutz (0.95)	
forbearance (0.92)	abstinence (0.84), sobriety (0.95), **temperance (0.89)**	Enthaltsamkeit (0.86), **Mässigung (0.82)**, Nüchternheit (0.68)	Einsichtigkeit (0.27)
	economy (0.74), **thrift (0.98)**	Sparsamkeit (0.87), **Wirtschaftlichkeit (0.87)**	Kalkulation (0.83), Masshalten (0.91), Wert (0.35)

(*continued on next page*)

Table 1 (*continued*)

Language-typical elements English	English values with German translations	German values with English translations	Language-typical elements German
credibility (0.98), dependability (0.89)	reliability (0.89), reliableness (0.80), **trustworthiness (0.93)**	Verlässlichkeit (1.0), **Vertrauenswürdigkeit (0.82)**, Zuverlässigkeit (1.0)	Konsistenz (0.50), Korrektheit (0.79), Pflichtbewusstsein (0.93), Pünktlichkeit (0.82)
service (0.41)	expediency (0.36), feasibility (0.83), practicability (0.83), pragmatism (0.83), usability (0.98), **usefulness (0.95)**	Brauchbarkeit (0.95), Machbarkeit (0.81), Nutzen (0.76), **Nützlichkeit (0.95)**, Pragmatismus (0.81), Zweckmässigkeit (0.95)	
animation (0.96), buoyancy (0.57), spiritedness (0.98)	activity (0.5), alertness (0.83), energy (0.98), exuberance (0.96), liveliness (1.0), **vitality (0.98)**	Aktivität (0.95), Ausgelassenheit (0.59), Energie (0.95), **Lebendigkeit (1.0)**, Lebensfreude (0.64), Regsamkeit (0.77)	Dynamismus (1.0), Spannung (0.45)
awareness (0.85), insightfulness (1.0)	experience (0.88), gnosis (0.58), knowledge (1.0), maturity (0.63), **wisdom (1.0)**	Erfahrung (0.95), Erkenntnis (0.75), Reife (0.80), **Weisheit (0.90)**, Wissen (1.0)	Klugheit (0.70), Schlauheit (0.55)

For generating the value maps, the word bags of the groups (the union of all word bags of the group member terms) have been used to calculate the inter-group dissimilarities as outlined in Sections 2 and 3. Those contained between 22 and 219 terms in English (mean: 94, median: 85 terms) and 8 to 397 terms in German (mean: 166, median: 154 terms). The resulting value group maps differed between English and German, as expected. If the local neighborhoods within a map of a single language are investigated, we find in many cases plausible results. For example, close neighbors of the group "harmony" are "friendliness", "empathy" and "solidarity", neighbors of the group "heroism" are "courage", "aggressiveness" and "purposefulness". Other neighborhoods are implausible (e.g., the closeness of "prudence" and "beauty", or "love" and "aggressiveness") – but we remind the reader that a 2-dimensional representation can never completely match the real neighborhoods in a 78-dimensional space. That is, the maps only can serve as heuristics and "real" neighborhoods have to be checked in the original data. Taking the example of the value group "love", we find that in the original space the closest neighbors are "piety", "enthusiasm", "passion", "respect" and "loyalty", i.e. the closeness of "aggressiveness" is an artefact of the algorithm that optimizes the mutual positioning of all groups (as outlined previously, the smaller size of the "love" circle indicates a topological inadequacy).

However, in using the maps heuristically for a comparison across languages, we can identify equal values terms that seem to differ in their neighborhood, and we then can check the original data for the closest neighboring groups. We did this analysis for four value pairs for which we had indications that these values may be understood differently in Germany compared to the United States (Table 2). For example, according to US opinion, the semantic neighborhood of autonomy and prosperity in relation to America's achievement-oriented culture may show the "typical" American mentality of individualism (Spence, 1985), whereas (social) equality and responsibility are terms often used in the German political discourse (the website of the German *Bundeszentrale für politische Bildung*, check for the keyword "Soziale Gerechtigkeit"; see http://www.bpb.de).

Indeed, Table 2 indicates striking differences in the semantic neighborhoods of the value terms investigated. For example, the American understanding of equality goes into the direction of "treating everybody in the same way", whereas the German "Gleichheit" reflects the fact that equality in Germany is often discussed with respect to protecting the society from economic inequality in order to preserve harmony within society. In addition, being autonomous in the US involves mainly personal control over the environment, whereas the semantic neighborhood of the German "Autonomie" indicates a more relational understanding of autonomy. This analysis demonstrates the heuristic benefit of these types of value maps.

Table 2
Semantic neighborhoods of selected English–German value pairs: Displayed are the four closest value groups of either value (in case of autonomy in English, the fourth-closest croup consisted of two equidistant groups).

Value group	US-American understanding	German understanding
Autonomy [Autonomie]	Authority, Liberty, Professionalism, Purposefulness/Strength (equally close)	Freiheit [freedom], Grosszügigkeit [generosity], Offenheit [openness], Sicherheit [security]
Equality [Gleichheit]	Objectivity, Fairness, Integrity, Serenity	Rechtmässigkeit [lawfulness], Harmonie [harmony], Empathie [empathy], Solidarität [solidarity]
Prosperity [Wohlstand]	Nonmalefience, Contentment, Profit, Usefulness	Nützlichkeit [usefulness], Geschicklichkeit [skill], Perfektion [excellence], Leistungsfähigkeit [strength]
Responsibility [Verantwortung]	Commitment, Piety, Loyalty, Obedience	Loyalität [loyalty], Vertrauenswürdigkeit [trustworthiness], Stabilität [stability], Rechtmässigkeit [lawfulness]

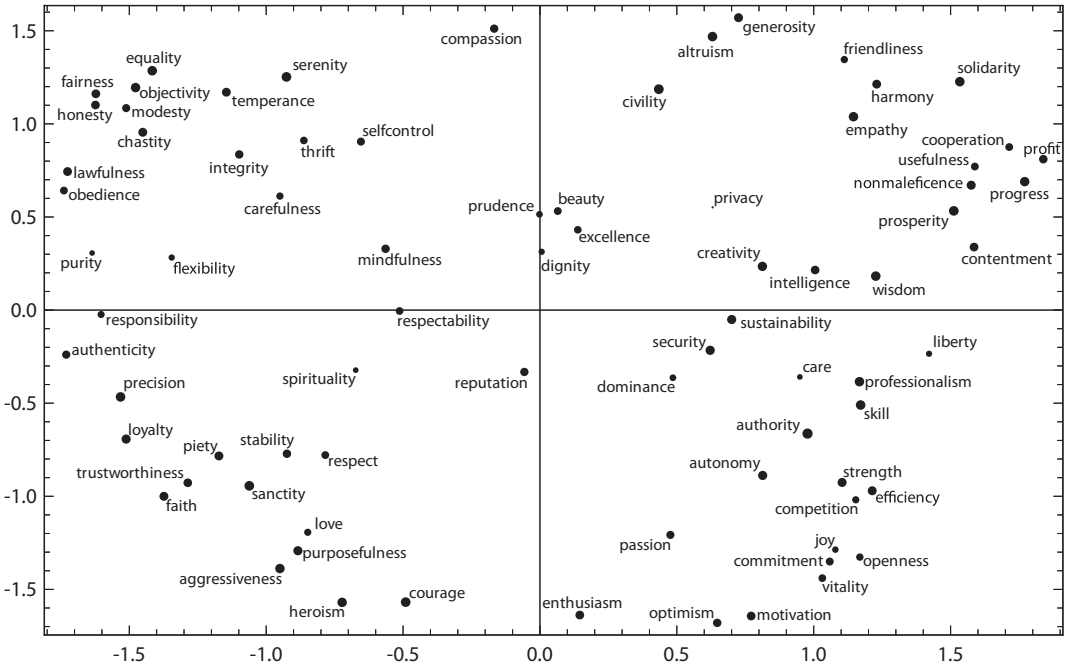

Fig. 2. Value map of the 78 English value groups. The point size reflects how well a value is positioned relative to all other values in the map compared to the dissimilarities in the original data space.

4.2. Survey study

In both the English and the German study, participant's ability to detect the distractor term served as quality check. Those participants who were unable to identify the distractor term were excluded from the sample.

In the English survey study, 449 persons provided data. Based on the quality check, 169 persons (37.6%) were excluded from the analysis due to their inability to detect the distractor term.[6] Out of 280, 143 (51.1%) participants were male, median age was 30 (range: 18–75). Each participant evaluated 12 synonym sets, so that we had in the mean 43 evaluations per value group.

In the German survey, 246 persons provided data and only 45 (18.3%) were excluded based on our quality criterion. Out of 201 participants, 81 (40.3%) were male, median age was 28 (range: 20–71). We can only speculate about the reasons of the much lower exclusion rate among the Swiss participants (18.3%) compared to the English sample (37.6%). We suspect that the lower exclusion rate is mainly an effect of the different sampling methods.[7] Furthermore, as the participants in the German survey

[6] We also checked whether the distractor terms used really have been reliably identified as distractors, and we found that in 12 groups (German: 5 groups) other terms were at least as often deleted as the distractors (usually terms that are uncommon in everyday language as, e.g., "mettle" in English or "Kaltschnäuzigkeit" in German) – those groups were not used in the quality check.

[7] In the US-survey, we used *Amazon Mechanical Turk* for participant recruitment, where each person is paid a small amount of money for survey completion. This may have been an incentive for fast survey completion, which may have increased the likelihood for errors when identifying distractors. In fact, the German participants worked longer on the tasks used almost 2 min per group evaluation, compared to about 1 min in the English survey.

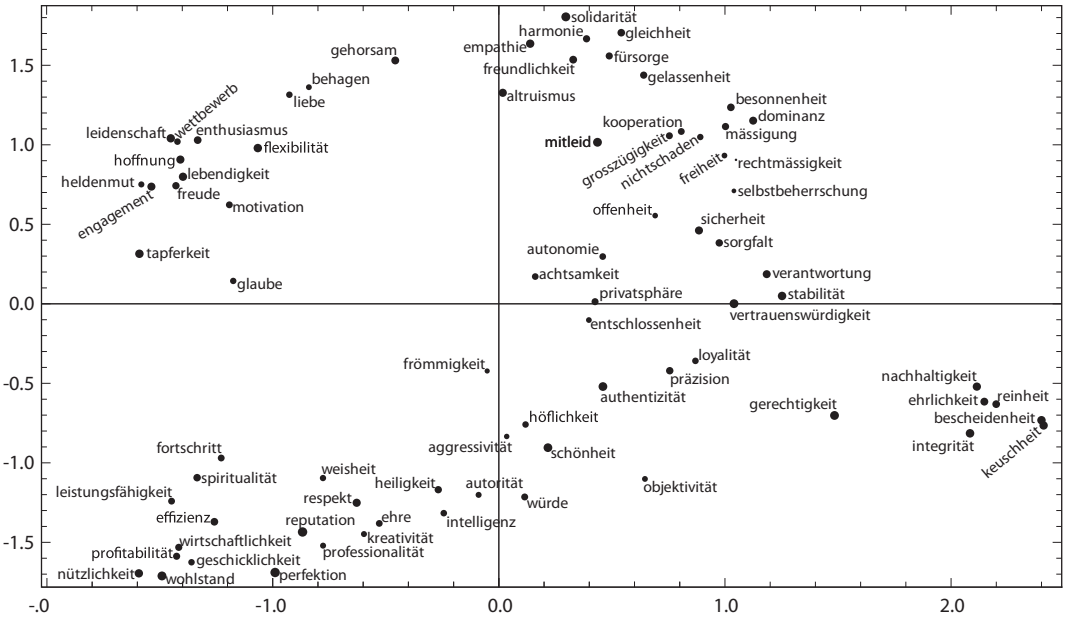

Fig. 3. Value map of the 78 German value groups. The point size reflects how well a value is positioned relative to all other values in the map compared to the dissimilarities in the original data space.

only evaluated 10 groups per person and not all of them evaluated all groups, the number of evaluations per value group was only 22 (compared to 43 in the English survey).

The relative frequency with which the corresponding value term was deleted from the groups is included in Table 1 (number in brackets after each value term). About two thirds (English: 77.4%, German: 71.0%) of the terms were kept in the groups in at least 75% of the cases (=possibilities to delete a term) and only about one tenth (English: 10.4%, German: 9.6%) of the terms were deleted in at least 50% of the cases. The latter consisted of two different types of terms: either terms that were rarely used in the everyday language and thus have been deleted due to ignorance (e.g., "mettle" in English, "Kaltschnäuzigkeit" in German), or terms for which real disagreement was present whether the term should be an element of the value group or not.

We also checked for a correlation between value group stability as estimated in the thesaurus study and evaluated in the survey. In the thesaurus study, we used the mean of the pairwise overlap of the word bags of all group members as indicator of group stability. In the survey study, we used the mean of the pairwise relative frequency that two group members have not been deleted by the same person as indicator of group stability. We found a Pearson correlation coefficient of 0.30 ($p = 0.007$) for the English value groups and a correlation coefficient of 0.40 ($p < 0.001$) for the German value groups, which is only a moderate positive relationship.

5. Discussion and conclusion

In this project, we have presented a methodology for investigating value pluralism in terms of semantic similarity of value terms. Our approach is inspired by the psycholexical tradition and uses a novel data source (thesaurus databases) and visualization tool in order to map value spaces. Our process was iterative, starting with actual language use as recorded by thesauruses, then applying statistical analyses whose results were then reviewed by actual human beings. In this way we employed a multimethod approach to drawing value maps.

We followed a bottom-up approach by starting from a large number of value terms instead of pre-selecting a low number of "core values" that then become the object of empirical investigation. By adding a survey study, we were able to demonstrate that the value groups identified using the thesaurus approach correlated with the similarity assessment of participants.

We believe that the results help to solve several problems in empirical value research:

(1) Our approach may be useful for empirical studies designed to examine values based on questionnaire items or vignettes. Often, it is necessary to include some variability in the wording of these textual elements, e.g., in order to hide what a scale is intended to measure or to increase the ecological validity of vignettes by avoiding repetition of terms. The thesaurus approach allows quantifying the similarity of synonyms and the choice of suitable replacements when constructing vignettes and/or items. The terms and data listed in Table 1 thus provide a practical instrument for researchers.
(2) In order to understand value pluralism, it is important to perform empirical studies in different cultural settings. This requires translation of instruments that

include value terms, and it is well known that such translations are methodologically challenging (Davidov & De Beuckelaer, 2010). As our analysis demonstrates, even with correct translation, there are differences in the semantic neighborhood of value terms – and such differences could point to hidden difficulties when performing such translations. For example, a study that investigates equality in different countries may need to take into account that the semantic neighborhood of this term is rather different in the German compared to the USA context. The fact that thesaurus databases are the result of language practice makes such a claim plausible. Therefore, the value maps presented could point to such difficulties in case of intercultural studies.

(3) Furthermore, we believe that the methodology presented here could become a heuristic for investigating value pluralism as such. One could, for example, hypothesize that values that are "close" together are less likely to be in conflict compared to values that are further away from one another (the Schwartz value map is constructed in a similar way).

We add that the empirical survey points to some difficulties when using *Amazon Mechanical Turk* for recruiting participants. The exclusion rates based on our tests were twice as high compared to standard internet recruitment, which seems to be related to a much faster (and probably more careless) survey completion. We therefore believe that the integration of test questions is indispensable when using *Amazon Mechanical Turk* for recruiting participants (see also Paolacci, Chandler, & Ipeirotis, 2010).

We remind the reader of several important shortcomings of our study: First, despite the large number of initial value terms, there are likely still important terms that have been missed and that fit with our broad definition of "value". However, the methodology allows for an easy inclusion of more value terms. This would require to generate the word bag of the new candidate terms and to investigate, whether they share a strong affinity to an existing group or not. Depending on the result, they may be integrated into a group or be treated as an independent group and then used as input into the visualization algorithm. Adding single terms does not require reproducing the complete procedure outlined in Fig. 1. Rather, the word bag of the new term can be extracted from a standard Thesaurus database and then be used for calculating the word bag overlaps (the Supplementary information contains the word bags of all 460 German and English values used in our study). Certainly, it is also possible to split existing groups and treat the parts as independent group, depending on the problem under investigation. Second, the nature of the data does not allow deciding, whether the similarities of values found only reflect purely pragmatic use of words within (American) English and German, or whether they actually represent relations between mental concepts. For example, a potential hypothesis one could have is that the differences in semantic neighborhoods of terms point to differences in the accessibility of the related mental concepts (Higgins, 1996). It could be that Germans, when confronted with the value "equality", are much more likely to activate the concept of "harmony" compared to Americans. This requires additional empirical research that is beyond the scope of this work. Third, the similarity measure used may misguide the analysis as we did not take into account differences in usage frequency of terms. For example, two pairs of terms may share an overlapping set of the same relative size – but in one case the overlap may consist of terms that are very frequently used as synonyms for the terms under investigation, whereas in the other case, the word bags share synonyms that are only very rarely used. It would be plausible to treat these two cases differently – however, this would again need additional empirical data on how frequent certain terms are used as synonyms. To diminish this problem, services like *Google Ngram* (https://books.google.com/ngrams) could be used for obtaining proxy data regarding term frequency in synonymy relations. Fourth, our current methodology involved expert feedback for group construction. The number of experts could be considered to be too low for that respect. However, we remind the reader that we included this step for heuristic reasons to decrease the number of values and to allow for visualizations that are easier to understand. One may skip this part to the cost that the resulting maps are harder to interpret. Fifth, the survey study may be considered insufficiently representative for matching the general language practice regarding similarity assessment. This is more problematic for the German survey that approached Swiss-German speaking people, i.e. a rather small sample within the German-speaking world. Furthermore, *Amazon Mechanical Turk* is known to have a higher demographic diversity compared to university student populations, although some differences to the general population remain (Ross, Irani, Silberman, Zaldivar, & Tomlinson, 2010). Sixth, the differences found may also reflect to some degree differences of the thesaurus databases themselves. Generally, linguistic reference books may insufficiently reflect the social and political contexts of language usage. Furthermore, dictionaries and thesauri rarely openly discuss why some lexical items are included and others are not. Remind that the German database generated in the mean more synonyms per value term compared to the English database. Thus, a deeper analysis of the history and curation of these databases may be required, which was, however, beyond the scope of this study. Finally, it might be tempting to draw general conclusions about human values from our work, but it must be said that value maps are influenced not only by a particular culture or subculture, but also by the implicit assumptions of a dominant worldview. For example, the current dominant worldview in several countries (e.g., detached individualism, selfishness and low social support) were unheard of and a death knell in most societies for the vast majority of human genus existence (Ingold, 1999; Sahlins, 2008). Thus

we can only conclude that what we have measured is a place in time. We cannot draw any conclusions about human nature nor tie the results to genetic or evolutionary history.

In summary, we have proposed to use thesaurus databases as a novel instrument for analyzing the interrelation of values (and other concepts). We show how machine classification and expert feedback can interact for creating a meaningful output, and we have implemented a novel visualization technique that aligns with the bottom-up nature of our approach. The survey study confirmed that the similarities found in the thesaurus analysis match to some extent with general language understanding. The value maps created in this way as well as the data regarding value similarity will help, to our understanding, to improve instruments in intercultural value research and serve as a heuristic for better understanding value plurality.

Acknowledgements

We thank the student members of the Moral Psychology Lab of Darcia Narvaez (E. Murray, A. Fish, O. Godby, Z. Kacprowicz, M.C. Larme, A. Lawrence, E. Peterson, E. Roof, C. Thompson, J. Young) for their support in generating the value lists. Furthermore, we thank our external experts (among others: M. Alfano, J.-C. Heilinger, R. Baumann-Hölzle, L. Huber, C. Ineichen, K. Ladd, S. Müller, A. Naves de Brito, M. Waldvogel) for their contribution to this work.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cogsys.2016.02.003.

References

Allport, G., & Odbert, H. (1936). Trait-names: A psycho-lexical study. Princeton, NJ: Psychological Review Company.

Bilsky, W., Janik, M., & Schwartz, S. H. (2011). The structural organization of human values. Evidence from three rounds of the European Social Survey (ESS). *Journal of Cross-Cultural Psychology, 42*(5), 759–776.

Blatt, M., Wiseman, S., & Domany, E. (1996). Superparamagnetic clustering of data. *Physical Review Letters, 76*, 3251–3254.

Davidov, E., & De Beuckelaer, A. (2010). How harmful are survey translations? A test with Schwartz's human values instrument. *International Journal of Public Opinion Research, 22*(4), 485–510.

Davidov, E., Schmidt, P., & Schwartz, S. H. (2008). Bringing values back in the adequacy of the European social survey to measure values in 20 countries. *Public Opinion Quarterly, 72*(3), 420–445.

Eisenberg, N. (2000). Empathy and sympathy. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 677–691). New York, NY: Guilford.

Galton, F. (1884). Measurement of character. *Fortnightly Review, 36*, 179–185.

Haidt, J., & Joseph, C. (2007). The moral mind: How 5 sets of innate moral intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.). *The innate mind* (Vol. 3, pp. 367–391). Oxford: Oxford University Press.

Hare, R. M. (1954). Universalisability. *Proceedings of the Aristotelian Society, 55*, 295–312.

Harris, R. (1973). Synonymy and linguistic analysis. Toronto: University of Toronto Press.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: Guilford.

Hüllen, W. (2004). A history of Roget's thesaurus. Oxford: Oxford University Press.

Huntington, S. P. (1993). The clash of civilizations? *Foreign Affairs, 72*(3), 22–49.

Ingold, T. (1999). On the social relations of the hunter–gatherer band. In R. B. Lee & R. Daly (Eds.), *The Cambridge encyclopedia of hunters and gatherers* (pp. 399–410). New York: Cambridge University Press.

Jones, S. K. (1986). Synonymy and semantic classification. Edinburgh: Edinburgh University Press.

Ott, T., Eggel, T., & Christen, M. (2014). Generating low-dimensional denoised representations of nonlinear data with superparamagnetic agents. In *Proceedings of the 2014 international symposium on nonlinear theory and its applications (NOLTA)*, Lucerne, Switzerland, September 14–18 (pp. 180–183).

Ott, T., Kern, A., Steeb, W.-H., & Stoop, R. (2005). Sequential clustering: Tracking down the most natural clusters. *Journal of Statistical Mechanics: Theory and Experiment*, P11014.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*(5), 411–419.

Pearce, B., & Littlejohn, S. W. (1997). Moral conflict: When social worlds collide. Thousand Oaks, CA: Sage Inc.

Quine, W. V. O. (1960). Word and object. Cambridge, MA: MIT Press.

Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review, 60*, 20–43.

Quintelier, K., & Fessler, D. T. (2012). Varying versions of moral relativism: The philosophy and psychology of normative relativism. *Biology and Philosophy, 27*(1), 95–113.

Rokeach, M. (1973). The nature of human values. New York: The Free Press.

Ross, J., Irani, I., Silberman, M., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers? Shifting demographics in Amazon Mechanical Turk. In *Proceedings of the conference on human factors in computing systems*, Atlanta, GA, April 10–15 (pp. 2863–2872).

Sahlins, M. (2008). The Western illusion of human nature. Chicago: Prickly Pear Paradigm Press.

Schwartz, S. H. (2012). An overview of the Schwartz theory of basic values. *Online Readings in Psychology and Culture, 2*(1). http://dx.doi.org/10.9707/2307-0919.1116.

Schwartz, S. H. (2006). A theory of cultural value orientations: Explication and applications. *Comparative Sociology, 5*, 136–182.

Schwartz, S. H., & Sagie, G. (2000). Value consensus and importance: A cross-national study. *Journal of Cross-Cultural Psychology, 31*(4), 465–497.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theory and empirical tests in 20 countries. In M. Zanna (Ed.). *Advances in experimental social psychology* (Vol. 25, pp. 1–65). New York: Academic Press.

Shweder, R. A. (1993). The cultural psychology of emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 417–431). New York: Guilford.

Spence, J. T. (1985). Achievement American style: The rewards and costs of individualism. *American Psychologist, 40*(12), 1285–1295.

Taylor, P. W. (1978). On taking the moral point of view. *Midwest Studies in Philosophy, 3*(1), 35–61.

Trier, J. (1931). *Der deutsche Wortschatz im Sinnbezirk des Verstandes: Die Geschichte eines sprachlichen Feldes. Von den Anfängen bis zum Beginn des 13. Jahrhunderts* (Vol. 1). Heidelberg: Quelle und Meyer.