Primenjena psihologija
Vol. 16, No. 3, pp. 403-434, 2023

Research Article

The development and preliminary validation of the Serbian value lexicon – An archival approach to value measurement

Zoran Pavlović [1] and Tijana Vesić Pavlović [2]

[1] *Department of Psychology, Faculty of Philosophy, University of Belgrade*

[2] *Laboratory for Foreign Languages, Faculty of Mechanical Engineering, University of Belgrade*

ABSTRACT

Values refer to stable beliefs and principles held by individuals, which guide their attitudes, behaviours, and judgments, and play a crucial role in shaping their identities and interactions with others. Studying values in social psychology is important as it provides insights into the motivational forces that drive individuals' behaviour and decision-making, shaping the dynamics of interpersonal relationships and societal interactions. The aim of this paper is to test the possibility of measuring basic values in the archive and text materials. Based on the Schwartz's theory of values and earlier work on the value lexicon in English, the Serbian lexicon of values was developed and preliminarily validated on a large-scale Internet-based survey. The lexical co-occurrence of words in the natural language use on the Internet was analysed in order to assess the convergent, discriminant and predictive validity of the lexicon. Lexical co-occurrence analysis showed that the words representing the same values co-occurred significantly more in comparison to the words denoting different values. The pattern of correlations between the values measured in the archive material on the Internet using the value lexicon showed high convergence with the pattern of correlations between the values assessed by the self-reported measures used in the European Social Survey in 2018. The relative prominence of the specific values on the official websites of the exemplar societal institutions

and organizations identified by the value lexicon was in line with the expectations and preliminarily confirmed the criterion validity of the lexicon of values. Possible applications of the lexicon of values, as well as some methodological issues pertaining to its future use, are discussed in the final part.

Key words: basic values, value lexicon, archival research, Internet-based research, Serbia

UDK: 159.9.072(497.11)
DOI: 10.19090/pp.v16i3.2465
Received: 04.04.2023.
Revised: 03.07.2023.
Accepted: 25.08.2023.

✉ Corresponding authors' email: zoran.pavlovic@f.bg.ac.rs

Introduction

Values research is a critical area of study in social psychology that helps to understand the guiding principles and beliefs that shape individuals and societies. By examining the values that people hold, it is possible to gain insights into how these values influence behaviour, decision-making, attitudes, and social relationships (e.g., Roccas & Sagiv, 2017; Rokeach, 1973). This knowledge is particularly relevant in today's diverse and interconnected world, where cultural differences and clashes can lead to misunderstandings and conflicts. Moreover, values research can inform interventions and policies aimed at promoting social harmony, justice, and well-being by identifying and addressing the underlying values that drive behaviour (Crompton, 2010; Pavlović, 2021). Overall, values research plays a vital role in advancing our understanding of human behaviour and improving social outcomes.

Values are, in (social) psychology, typically studied via self-report measures and self-administered questionnaires that vary in the number of items, i.e., the measured values, stimulus types (e.g., words versus visual materials), response format (e.g., ranking versus rating) etc. (for an extensive overview see Roccas, Sagiv & Navon, 2017; Pavlović, 2021). But, simply put, all of these imply administering a questionnaire and relying on the self-report measures, which have some well-known weaknesses, such as social desirability, consistency, and memory biases (Araujo et al., 2017; Bardi et al., 2008; Podsakof et al., 2003; Roccas et al., 2017). Furthermore, impression management and self-deception, acquiescent responding, extreme responding, pattern and random responding etc. are well-known "troubles" of the self-report method in general (Paulhus & Vazire, 2007). Also, the need to develop the measures of individual values in the situations where reliance on questionnaires is "undesirable, impractical, or impossible" (Bardi et al., 2008, p. 483) is often emphasised. The measurement of individual values based on, broadly speaking, archive materials, primarily the text, user-generated or natural language use data, has been gaining much attention recently.

The natural language use approach in the study of values examines the individuals' language patterns, such as their words, expressions, and discourse, to gain insights into their personal values and belief systems. By analysing the linguistic cues and semantic structures employed by individuals, researchers can uncover the underlying value orientations and understand how they shape individuals' thoughts, attitudes, and behaviours (Pennebaker et al., 2003; Ponizovkiy et al., 2020). Similarly, by examining the linguistic cues and word choices employed by individuals, researchers can uncover meaningful associations between language use and personality dimensions (Pennebaker & Stone, 2003; Schwartz et al., 2013). Such procedures do not just overcome the self-report measures' weaknesses but offer a possibility to analyse values from inaccessible populations and in a great variety of time and places. Furthermore, values research simply cannot ignore the large amount of textual data produced and shared online in any moment (Ponizovkiy et al., 2020).

The basic assumptions of the natural language use approach are in line with the well-known psycho-lexical approach to the study of personality, which focuses on capturing the rich diversity of human personality traits by analysing and categorizing the natural language descriptors (Goldberg, 1981; Sausier & Goldberg, 1996). It seeks to identify and organize the fundamental dimensions of personality through the analysis of the words and language used to describe individuals' characteristics and behaviours. Although mostly used in personality psychology, it has already been successfully tested in the study of values as well (De Raad et al., 2016).

Relying on the Schwartz's theory of personal values (Schwartz, 1992; Schwartz et al., 2012) and earlier successful attempts at developing the lexicon of values (Bardi et al., 2008), this paper aims to develop the archival approach to the study of values and value lexicon in the Serbian language. We proceed as follows. We first shortly present the Schwartz's value theory and the archival approach to the study of values developed by Bardi et al. (2008). After that, we describe the current study and applied methodology in more detail. The obtained results are then shown and discussed. We conclude

with a number of possible implications and the future usage of the developed value lexicon in Serbian.

Schwartz's value theory and the archival approach to the study of values

The idea of measuring values in archive material is as old as the modern study of values and goes back to Rokeach's (1973) seminal study on the nature of human values. Arguing that the crucial difference between political ideologies lies in their focus on values of freedom and equality, Rokeach demonstrated this argument by content analysis of the documents written by the proponents of the, then, four main political ideologies: fascism, socialism, communism, and capitalism. Still, the main impetus for the archival study of values came from the attempts to apply the influential Schwartz's value theory to different types of textual data.

Schwartz's theory of values (Schwartz, 1992; Schwartz et al., 2012; Schwartz & Cieciuch, 2022) is a comprehensive framework for understanding and categorizing the values that individuals and societies prioritize. Values are defined as "trans-situational goals, varying in importance, that serve as guiding principles in the life of a person or other social entity" (Schwartz, 1994, p. 21). The theory proposes that there are ten basic values that are universal across cultures and time periods, and that these values can be organized into a circular structure based on their motivational compatibility and opposition to each other. These ten values include (Schwartz, 1992): *Power* (control and influence over others), *Achievement* (personal success and accomplishment), *Hedonism* (pleasure and enjoyment), *Stimulation* (new experiences and excitement), *Self-direction* (independence and creativity), *Universalism* (social justice, equality, and tolerance), *Benevolence* (helping others and promoting their welfare), *Tradition* (preserving and respecting cultural and religious traditions), *Conformity* (obeying rules and fitting in with others), and *Security* (safety, stability, and order). This typology of values is based on the idea that values in the form of conscious goals represent three universal human existence requirements: 1) the necessity of satisfying biological needs,

2) the coordination of social interactions and 3) the survival and functioning of groups (Schwartz 1992; 1994; Schwartz et al., 2012; Schwartz & Cieciuch, 2022).

The circular structure of values indicates that some values are compatible and reinforce each other, while others are in opposition and can create tension (Schwartz 1992; Schwartz et al., 2012; Schwartz & Cieciuch, 2022). For example, universalism and benevolence values are compatible with each other, while power and benevolence values are in opposition. Schwartz has also identified several higher-order values that represent overarching goals, such as self-transcendence (combining universalism and benevolence values) and self-enhancement (combining power, achievement, and hedonism values) (Schwartz 1992; 1994; Schwartz et al., 2012; Schwartz & Cieciuch, 2022).

Schwartz's theory of values has been applied in various fields to understand the motivations and behaviours of individuals and groups. For example, it has been used to explain prosocial behaviour (Sanderson & McQuilkin, 2017), political activism (Veccione et al., 2015), and organizational behaviour (Arieli & Tenne-Gazit, 2017). The theory suggests that values are a key driver of human behaviour, and that understanding individual's values can provide insight into their decision-making processes and behaviour patterns.

Several self-report instruments for the measurement of Schwartz's basic values have been developed, most notably Schwartz's Value Survey (SVS, Schwartz, 1992), the 21-item Portrait Values Questionnaire (PVQ-21, Davidov et al., 2008) and the recently revised Portrait Values Questionnaire (PVQ-RR, Schwartz et al., 2012; Schwartz & Cieciuch, 2022). In recent years, the theory has been increasingly used in the 'questionnaire-free', theory-driven archive analyses. One of the earliest studies of this kind was conducted by Schwartz and Ros (1995), who contrasted the Western Europe and the United States cultural ideas, as expressed in the motto of the French Revolution and the American Declaration of Independence. Autonomy values were shown to be shared, but differences were visible in the importance given to Egalitarianism, Harmony, and Mastery values. Similarly, one study content-

analysed the values expressed in Czars' and Presidents' speeches in Finland from 1809 to 2000 (Portnam, 2014), using the qualitative coding manual created based on Schwartz's theory. In addition to the ten basic values, two value types, *Spirituality* and *Work-related* values, were identified.

Several studies used the Linguistic Inquiry and Word Count (LIWC, Pennebaker et al., 2007), the most commonly used language analysis tool for investigating the relation between word use and psychological variables, to analyse the user-generated content on social networking sites, such as Reddit (Chen et al., 2014), Facebook (Mukta et al., 2010) and Weibo (Sun et al., 2014) and to (successfully) predict the PVQ scores from users' status updates. For example, in the study by Chen et al. (2014), for each Reddit user, one LIWC measure for each LIWC category based on the user's posts/comments was calculated. The frequency of word use in one LIWC category was then correlated with the PVQ scores and it was shown that, for example, the higher importance of the Hedonism values was related to lower anxiety, fewer prepositions and words about the family, but more swear words. Mukta et al. (2010) used the PVQ scores of 397 Facebook users as a dependent variable, while the analytically determined best subset of the LIWC categories was used as an independent variable. For example, the family, affect, anxiety, feel, and such LIWC categories of words could have been used for computing the self-transcendence values from users' statuses. Fifteen different subsets of the LIWC features were then used to predict the PVQ scores in a series of regression models; between 13% and 21% of the variance in the PVQ measures could be explained by the LIWC categories.

Bardi et al. (2008) began with the idea that natural language use was a reliable and accurate indication of psychological constructs and created a lexicon of value terms to represent Schwartz's (1992) individual-level value theory. This value lexicon was subjected to convergent, discriminant, and predictive validity tests. They demonstrated the lexicon's convergent and discriminant validity by analysing its convergence with the structure of individual values evaluated by standard self-report techniques using an Internet-based methodology. Moreover, the predictive validity of the value

lexicon was evaluated by correlating the patterns of individual values as depicted in the American newspapers with the objective indicators of behaviours during the course of the 20th century. For example, more prominent Achievement values were positively correlated with the yearly number of patented inventions, Hedonism was correlated with alcohol consumption per capita and Stimulation with the number of the movies released. Similarly, higher voting participation was related to more prominent Self-direction and unwed births with the Conformity value (reversely scored) (Bardi et al., 2008).

Most recently, Ponizovskiy et al. (2020) conducted a similar analysis of a much wider scope on the corpus of the social media posts, essays, and works of literary fiction, containing 525,901,609 words authored by 182,197 individuals. They developed a Personal Values Dictionary (PVD), which comprises more than 1000 value-laden terms for detecting the references to the personal values in text, demonstrating its content, construct, and criterion validity.

All these studies have shown the general usefulness of the theory-driven approaches[1] to the study of values in archive materials and the applicability of Schwartz's theory. Still, all these were developed and tested in English and similar studies in other languages are absent.

The Current Study

This study aims at developing and validating the value lexicon in the Serbian language. Schwartz's value theory and different questionnaires in Serbian had already been validated, are available and were used in a number of research studies (e.g., Marušić-Jablanović, 2018; Radović, 2010; Radović et al., 2019; Pavlović, 2021; Pavlović & Stepanović Ilić, 2021; 2022; Lazić et al., 2021). There were also previous studies that analysed the conceptualization of

[1] Another line of research, a data-driven or exploratory approach, similarly relies on the analysis of the user-generated and/or textual data, but with the main difference in the assumption that the content and dimensionality of specific values are to be empirically determined (see, e.g., Ponizovskiy et al., 2020).

values in the public discourse from the cognitive-linguistic perspective using newspaper materials (Vesić Pavlović, 2013). Still, the theory-driven archival approach based on Schwartz's theory, to the best of our knowledge, has never been used to study values in Serbian. Therefore, in the present study, the value lexicon representing ten personal values developed and validated by Bardi et al. (2008) served as a starting point; it was translated into Serbian, adapted, and then used in an Internet-based study.

The underlying assumption of this approach (in line with the psycho-lexical paradigm) is that natural language use provides a reliable and valid indicator of basic personality, cognitive, and social processes (Pennebaker & King, 1999; Pennebaker, Mehl, & Niederhoffer, 2003), and that, for any given construct, a lexicon of words that is indicative of the construct can be developed (Bardi et al., 2008). The degree to which words converge on the activation of the construct, their associative strength, can be measured by the relative co-occurrence of words (Bardi et al., 2008). Using Internet browsing for such a purpose is further justified by the studies that show the correlations between the number of Google hits and language norms regarding the word frequency estimates. For example, Blair et al. (2002), showed that Internet searches were a very useful instrument and an adequate indicator of the word frequency estimates. An archival approach which relies on this methodology thus seems reasonable and empirically supported by previous research, but solely in the English language. This study is hence an attempt to replicate and validate it in Serbian.

All said, the aims of this study are threefold: (1) to analyse the patterns of lexical co-occurrence of the words representing the same and different values, using the Internet as a source of archival data; (2) to compare the structure of the relationship between the values obtained in archive data and those from the self-report survey; and (3) to preliminarily test the criterion validity of the value lexicon, i.e. to demonstrate that it can identify the specific values in the type of text for which it can be reasonably expected to place special emphasis on them.

Method

The Development of the Value Lexicon

As stated, the development of the value lexicon started from the aforementioned study by Bardi et al. (2008). The list of words (30 of them in total) representing individual values, which was validated in that study, was first translated and back-translated by two independent researchers, following the same criteria used in the development of the English lexicon: each of the ten Schwartz's values was represented by three single words; the value labels itself (e.g., *sigurnost* [*security*]) was used as a word representing this value, when meaningful); only nouns were used; evidently polysemous words (having multiple meanings depending on the context) were avoided; in the situations where different alternative translations were possible, the words were chosen for the individual values that conveyed a certain meaning when they appeared together, although individually they may not inevitably indicate a specific value (Bardi et al., 2008).

The Refinement of the Value Lexicon

The translation process was monitored and evaluated by an independent researcher, an expert in value surveys. The lexicon words' translation was evaluated and compared to the existing translations of different versions of the questionnaires for measuring Schwartz's personal values that already exist in the Serbian language. The SVS questionnaire (Schwartz, 1992; Radović, 2010), PVQ-21 (Davidov et al., 2008; used in ESS round 9 in Serbia in 2018, see also: Pavlović & Stepanović Ilić, 2021; 2022), PVQ-RR (Schwartz et al., 2012; Radović et al., 2019) and Short Schwartz's Value Survey (SSVS) questionnaire, a short ten-item form (Lindeman & Verkasalo, 2005; unpublished data in Serbian), were used as guides during the final stage of the lexicon development. The main aim of refinement was to ensure that the translation of the lexicon of values was consistent with the terminology that appears in the existing translations. The list of words that were included

in the final form of the Serbian lexicon of values is shown in Table 1 (representing words in English are taken from Bardi et al., 2008, p. 486).

Table 1
Personal values and their representing words in the value lexicon

Value	Representing words in Serbian	Representing words in English
Power	*moć, snaga, kontrola*	*power, strength, control*
Achievement	*postignuće, ambicioznost, uspešnost*	*achievement, ambition, success*
Hedonism	*luksuz, zadovoljstvo, uživanje*	*luxury, pleasure, delight*
Stimulation	*uzbuđenje, raznovrsnost, stimulacija*	*excitement, novelty, thrill*
Self-direction	*nezavisnost, samostalnost, sloboda*	*independence, freedom, liberty*
Universalism	*jedinstvo, pravda, jednakost*	*unity, justice, equality*
Benevolence	*dobrota, velikodušnost, milost*	*kindness, charity, mercy*
Tradition	*tradicija, običaji, poštovanje*	*tradition, custom, respect*
Conformity	*uzdržanost, uvažavanje, brižnost*	*restraint, regard, consideration*
Security	*bezbednost, sigurnost, zaštita*	*security, safety, protection*

Data and measures

Joint probabilities. After the development of the final version of the Serbian value lexicon, the next steps in the procedure from the original study (Bardi et al., 2008) were replicated. They included a set of Google searches. The lexicon of values adequately represents Schwartz's 10 values if the words represent each of the values differently in terms of their relative frequency of co-occurrence in natural language used on the Internet. Put simply, if a certain set of words adequately represents a specific value, the co-occurrence of that particular set of words (e.g., *power* and *strength* as the indicators of the Power value) should be relatively more frequent than the co-occurrence of the words representing different values (e.g., *power* and *kindness*).

In order to obtain the data needed for the joint probabilities and co-occurrence calculations, Google searches of individual words from the lexicon were first performed, and then all the possible pairs of words from the lexicon were searched, resulting in a total of 465 Google searches (30 individual words and 435 of their possible pairs). All Google searches were limited to the websites in the Serbian language, Latin script and search terms in the nominative case[2]. Possible combinations of words from the value lexicon were searched following the rules of the so-called Boolean Google search terms (e.g., *moć* AND *snaga*). The searches were partially automated; the URLs for all 465 searches were first created and deployed in Google searches performed by R software (R Core Team, 2022), using the *RCurl* (Temple Lang, 2022a) and *XML* (Temple Lang, 2022b) packages. The main data returned referred to the number of Google search results (i.e., 'hits').

In line with the original procedure, lexical co-occurrence of the pairs of words for the same and different values was expressed in terms of their joint probabilities (JP), which were calculated by dividing the number of Google search hits for a pair of words (PH) and the average number of 'hits' for individual words (SH). For example, JP for *moć* and *snaga* was calculated as follows (hits expressed in 1,000s, see Appendix, Table 1 for the representative part of the row co-occurrence matrix):

$$JP_{moć*snaga}=(moć*snaga_{PH})/[(moć_{SH}+ snaga_{SH})/2]=5,770/[(5,210+8,780)/2]=5,770/6,995=0.824$$

This resulted in 435 JP measures, one for each word-pair combination.

Convergent and discriminant validity. To test the convergent and discriminant validity of the value lexicon, Bardi et al. (2008) first studied the prevalence of words from the lexicon in the web/electronic newspapers in the USA archive in the period from 1900 to 2000. The relative number of

[2] We will address the Serbian language peculiarities in more detail in the Discussion.

newspaper pages in which all three words for the individual value appeared (expressed as a ratio to the total number of pages) served as a measure of individual values. In this way, they obtained a yearly value score for each of the ten values during 101 years, which made it possible to calculate the intercorrelations between the values, and their convergence with the intercorrelations among the 10 values in participants' self-reports obtained in an independent survey.

Since a media/newspaper archive of a nearly similar volume in Serbian was not at our disposal (nor, to the best of our knowledge, exists), we once again relied on Google searches. We searched for the number of web pages containing all three words denoting each of the 10 values on a monthly basis, during the period from January 2012 to December 2022. This was done by limiting the web searches to specific time periods with the help of Google's advanced search option, i.e., the specification of a number of URLs used for Internet browsing[3]. More specifically, we determined the number of pages in the Serbian language on the Internet in which all three words representing, for example, Universalism (*jedinstvo* AND *pravda* AND *jednakost*) simultaneously appeared during the month of January 2012, and then during February 2012, and so on, for each month, ending with December 2022. The same procedure was repeated for each of the 10 values. This resulted in 132 monthly measures of the ten values, expressed as the number of the Google search hits in relative terms, i.e., the total number of hits for all three words (expressed in 1,000s) was divided by the number of hits for the word "*je*"[4]. These measures were then used to calculate the intercorrelations between values in the archive material.

[3] All other technicalities related to the Google search are identical to those described earlier.

[4] The form of the auxiliary verb "biti" (to be) was arbitrarily chosen, with the reasonable assumption that it was highly unlikely that there existed a webpage in Serbian without this word. With this, we wanted to imitate the procedure that Bardi et al. (2008) had used, treating the number of newspaper pages containing the article "the" as a baseline.

As a self-report measure for calculating intercorrelations between the 10 values, the European Social Survey data (ESS), round 9, conducted in Serbia in 2018, were used (European Social Survey European Research Infrastructure, 2021). This study was conducted on the nationally representative sample of the Serbian citizens (N=2,043) and Schwartz's personal values were measured using the PVQ21. The measures of individual values are known to show rather lower internal reliability, since they are measured by a small number of items, selected to cover the different conceptual components of the value (e.g. Schwartz, 2003); our study showed that α for the ten values ranged from .44 (Conformity) to .78 (Hedonism). Multidimensional scaling (PROXSCAL) on 10 computed basic values (S-stress=.002; DAF=.99, Tucker's *phi* = .99) showed the satisfactory fit with the circular structure and ordering of values[5].

Two sets of 55 unique intercorrelations between the ten values were computed and then compared (as in the original procedure, the obtained correlation coefficients were Fisher Z-transformed before the analysis).

Predictive/criterion validity. To test the predictive validity of the value lexicon, Bardi et al. (2008) correlated the yearly measure of values, expressed in the American newspaper pages, with the yearly measures of different patterns of collective behaviour that could be treated as value-expressive behaviours. For example, the yearly alcohol consumption per capita was treated as an objective indicator of behaviour that expresses the value of Hedonism and was indeed correlated with the value of Hedonism measured in the newspapers. Due to a far narrower time frame observed in our study and the lack of relevant societal data that could be treated as value-expressive behaviour on a monthly basis (or significantly vary on a monthly basis at all), we deviated from the original procedure in this respect. We

[5] The full analysis report is available at https://osf.io/nk8z7/?view_only=3c2327fb3db54f1c9b51e984cfdd15cb.

applied a slightly different logic by assuming that the predictive validity of the value lexicon can be demonstrated in a sort of a test-standard scenario.

Certain institutions and organizations within the Serbian society can be perceived as bearers or promoters of specific values. For example, it can be reasonably expected that the Serbian Orthodox Church, in terms of ten values, should promote the value of Tradition and/or Benevolence more than, say, the value of Hedonism or Achievement. With this in mind, we selected four specific institutions and organizations as some sort of illustrative examples or showcases of the lexicon criterion validity. These were: *the Serbian Orthodox Church* (with the expectation that, as a protector of religious and cultural heritage, it should primarily exemplify Tradition), *the Serbian Army* (bearing in mind its main purpose of securing the national safety, it should promote Security), *the Civic Initiatives* (being one of the most important and visible NGOs and strongly voicing human rights and freedoms, it was assumed it should promote the value of Self-direction) and *the Serbian Association of Managers* (the organization aimed at promoting business, entrepreneurship and leadership, which should transpose to the promotion of Power and Achievement[6]). The criterion validity of the value lexicon would be demonstrated by higher prominence of those values that should be more propagated by these institutions, as compared to other values from the model.

In order to test our assumptions, we applied the similar methodology as earlier described. The official websites of the afore-mentioned institutions (www.spc.rs, www.vs.rs, www.gradjanske.org, www.sam.org.rs) were searched for the combinations of the value words indicators representing each of the ten values, using the previously described Boolean search terms and "in site:" specification. Bearing in mind a rather limited amount of 'content'

[6] Bearing in mind the circular orderings of values in Schwartz's model, these expectations are only provisional since higher prominence of a certain value (e.g. Tradition) will probably be accompanied by relatively high prominence of adjacent values in the model (e.g. Benevolence or Conformity as adjacent to Tradition, or Power as adjacent to Achievement).

on these websites (as compared to site-unspecified searches), we introduced a slightly different procedure by searching for the co-occurrence of any of the value pairs for a specific value, and not all three words at the same time. For example, we searched for the number of 'hits' with *tradicija* AND *običaji*, then *tradicija* AND *poštovanje*, and, finally, *običaji* AND *poštovanje*; the number of search results was then averaged and divided by the number of 'hits' for the word "*je*", as a baseline. As such, each value was measured as an average proportion of the pages on a specific website that contained any word-pair combination serving as indicators of the specific values (for ease of inspection, we expressed it as a percentage of total pages).

All searches were carried out in the period January 25-27, 2023. The materials needed for the reproduction of the analysis reported here can be accessed at https://osf.io/nk8z7/?view_only=3c2327fb3db54f1c9b51e984cfdd15cb. The collected data were analysed by the IBM SPSS Statistics 21 software.

Results

Lexical Co-occurrence of Value Words in the Value Lexicon

For each of the ten values, Table 2 shows mean joint probabilities of the pairs of words representing the same value and mean joint probabilities of the pairs of words indicating different values (i.e., one word representing the target value and one word representing a different value). The mean joint probabilities for the co-occurrence of the terms indicating the same values should be greater than the mean joint probabilities for the pairings denoting different values, allowing the 10 values to be distinguished well by the value lexicon in the natural language use on the Internet. As expected, the mean joint probabilities for the co-occurrence of the pairs of words for the same value (M = .36) are significantly higher than the mean JP for the co-occurrence of the pairs of words for different values (M = .16), $F(1, 53) = 19.81$, $p < .001$, $\eta^2 = .27$, $d = .60$.

Table 2

Mean Joint Probabilities (JP) of Co-occurrence among the Words in the Value Lexicon

Value	Mean JP of the pairs of words representing the same value	Mean JP of the pairs of words representing different values
Power	.74	.24
Achievement	.40	.16
Hedonism	.21	.14
Stimulation	.13	.10
Self-direction	.42	.21
Universalism	.23	.15
Benevolence	.21	.12
Tradition	.44	.20
Conformity	.15	.12
Security	.66	.17

Convergent and Discriminant Validity of the Value Lexicon

Table 3 presents the intercorrelations between 10 personal values obtained in the archive data/natural language use on the Internet and in self-reported data, i.e., the ESS round 9 conducted back in 2018. Following the logic applied in the original study (Bardi et al., 2008), the Serbian value lexicon would serve as an indicator of individual-level values if the general patterns found in the 55 unique correlations acquired from the lexical co-occurrence of the 10 values in archive data displayed the same general patterns as 55 correlations gained from the general population of the Serbian citizens' self-reports of the 10 values. Similarly to the data reported in the original study (r = .93, p < .001), very high convergence between these two sets of correlations was registered, r = .94, p < .001[7].

[7] As in the original paper (Bardi et al., 2008), this analysis is based on the Fisher Z-transformed correlation coefficients.

Yet, the pattern of correlations between the values in the archive material is far from perfect. There are some surprisingly negative correlations between adjacent values, such as, for example, Hedonism and Achievement, or positive correlations between the opposite values, like Achievement and Benevolence. These relationships probably stem from various possibilities of the concurrent use of different and opposing value terms online. Online debates that present the conflicting values or concurrent use of antonyms (Bardi et al., 2008), as well as the inability to consider the valence of value with an automated Internet search, are but some examples. Put differently, the words denoting opposing values could have co-occurred, resulting in a positive correlation, although the context of their use was in fact meaningful (for example, "*zaštita bezbednosti* ne sme voditi ugrožavanju *samostalnosti* i *slobode*", denoting concurrently opposing to the value of Security and supporting the value of Self-direction).

Table 3

Intercorrelations between 10 values in the archive data and self-report data

	1	2	3	4	5	6	7	8	9	10
1. Power	1	*.21* **	*.04*	*.02*	*.00*	*-.43* **	*-.36* **	*-.33* **	*-.11* **	*-.18* **
2. Achievement	.42 **	1	*.04*	*.11*	*.08*	*-.35* **	*-.23* **	*-.38* **	*-.30* **	*-.22* **
3. Hedonism	-.57 **	-.30 **	1	*.43* **	*.18* **	*-.43* **	*-.31* **	*-.43* **	*-.49* **	*-.34* **
4. Stimulation	.31 **	.32 **	-.24	1	*.22* **	*-.39* **	*-.30* **	*-.46* **	*-.48* **	*-.37* **
5. Self-direction	.81 **	.24	-.41 **	.16	1	*-.21* **	*-.18* **	*-.39* **	*-.37* **	*-.30* **
6. Universalism	.19	.12	-.14	.16	.14	1	*.32* **	*.26* **	*.19* **	*.23* **
7. Benevolence	.53 **	.45 **	-.38 **	.38 **	.31 **	.34 **	1	*.26* **	*.06*	*.18* **
8. Tradition	.90 **	.40 **	-.54 **	.31 **	.80 **	.18	.51 **	1	*.39* **	*.18* **
9. Conformity	.40 **	.32 **	-.25 **	.39 **	.30 **	.18	.37 **	.40 **	1	*.12* **
10. Security	.86 **	.30 **	-.42 **	.11	.88 **	.12	.37 **	.83 **	.31 **	1

Note. ** p <.001 (the Bonferroni correction). Correlations below the diagonal are from the archive data (N = 132); correlations above the diagonal (given in italics) are based on the self-reported ESS round 9 data (N = 594).

Predictive Validity of the Value Lexicon

The percentage of pages on the official websites of four institutions and organizations that contained any same value word-pair combination are shown in Figure 1. Mostly in line with our expectations, a higher relative occurrence of specific values on the institutional websites was registered. Out of ten basic values, Tradition showed the highest prominence on the website of the Serbian Orthodox Church. Power and Security were by far the most frequent values on the website of the Serbian Army, while Self-direction was the most prominent on the Civic Initiatives webpage. The Serbian Association of Managers webpage did show a relatively higher occurrence of the Power and Achievement values, in comparison to other values, although the most prominent value was in fact Security.

The percentage of pages mentioning the Security value was generally high on all websites, except the Serbian Orthodox Church. Apart from indicating a generally high significance of the Security value, this suggests possible differences in the meaning of this value in specific cases. One can advocate the *securing* and *protecting* of human rights and freedoms (Civic Initiatives), on the one hand, or stress the necessities to deal with the issue of economic *security* (Serbian Association of Managers), on the other, in which case the words denoting one and the same value, Security, are used in quite a different context.

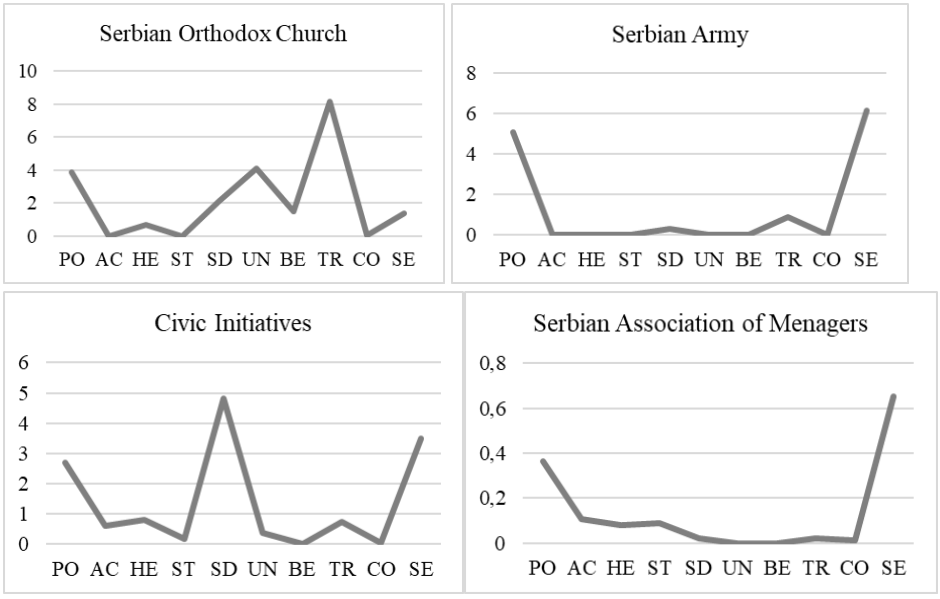

Figure 1. Relative occurrence of ten values on the official websites of four institutions expressed as a % of total pages.

Discussion and conclusion

The purpose of this research was to develop the value lexicon that would allow the measurement of value patterns without relying on self-report questionnaire responses, thereby facilitating the measurement of values over time and in real-world settings. The value lexicon was developed on the basis of Schwartz's (1992) value theory and recent advances in the study of associative meaning in natural language use and the existing value lexicon (Bardi et al., 2008). As such, the aims of this study were threefold. We, firstly, aimed to analyse the patterns of lexical co-occurrence of the words representing the same and different values, using the Internet as a source of archival data. The values were not assessed nor, as a matter of fact, expressed directly, but rather inferred from text, i.e., the co-occurrence of specific words in natural language use. It was shown that the lexical co-occurrence of the words representing the same values was higher than that of the words representing different values in natural language use on the Internet, i.e. the webpages in Serbian. Despite various external and uncontrolled influences

that could impact the way in which values are expressed in natural language, the significant increase in the co-occurrence of words related to the same values, as opposed to different values, across a vast amount of internet documents, highlights the value lexicon's ecological validity.

Furthermore, and related to our second aim, the pattern of correlations between the values measured on a monthly basis in the archive material on the Internet in the period from 2012 to 2022 showed a very high convergence with the self-report measure of values obtained in the ESS round 9, conducted on the nationally representative sample of the Serbian citizens. This demonstrated that the patterns of relationships between values expressed in natural language follow the theoretically described structure (Bardi et al., 2008; Schwartz, 1992). The expression of specific values in natural language seems to be guided by the very motivational dynamics described in the specific relationship of motivational (in)compatibility, which is in line with Schwartz's theory (Schwartz, 1992). Bearing in mind some deviations in the observed correlations between values, we must warn of certain inherent weaknesses of the co-occurrence measurement which does overcome typical weaknesses of the self-report measures but introduces some new ones as well. Yet, in the light of a large number of factors that could have influenced value expression in the language used on the Internet, which, importantly, could not have been controlled or even phantom, deviations are nothing but to be expected. Nevertheless, the significant convergence of the results from two distinctly different methodological approaches offers support for the validity of the value lexicon and for the indicators utilized to examine the possible relations between values and behaviours (Bardi et al., 2008).

Finally, the lexicon of values proved to be able to identify the differences in the prominence of different values on the official webpages of those institutions for which it can be reasonably expected to place special emphasis on the specific values from Schwartz's typology. Although only some kind of mimicking the values-behaviour analyses, if the website content can be viewed as a sort of collection of group behavioural traces, these data

indicate the existence of the meaningful values expression by the group or actual people in real time, despite it being affected by a number of external influences. Even more, it proved to be a useful venue for measuring the values of those 'agents' whose values are inaccessible to other types of value instruments, if any at all.

All said, the findings of this survey replicated the Bardi et al. (2008) findings in a different language and preliminarily validated the Serbian value lexicon and its utility for further use in the measurement of values at the time and in those places and contexts in which the use of a questionnaire is neither possible nor desirable. Available and relevant documents of individuals, groups, and organizations can be explored for the presence of values.

There are numerous potential applications of the lexicon of values. Studies in the Serbian language are far lagging behind both the theory- and/or data-driven archival approaches to the study of values. One reason for this is the lack of, in the broadest sense of this term, text corpora in Serbian that could be used for the present purposes. The other and probably more important reason for this is the lack of research that would offer a valid 'tool' for the measurement of values in this way or, at least, credible evidence that such a measurement in Serbian is possible at all. This study thus presents a small-scale attempt to preliminarily establish evidence of a kind that would, hopefully, motivate further and more thorough research. The developed Serbian value lexicon can be used for the purposes of analysing individual differences in values in user-generated content such as posts on social networking sites; further, organizational and group values can be measured by analysing their official documents, such as the content of the official webpages utilized here, or other relevant text materials, for example, party manifestos when addressing the issue of parties' value promotion (e.g. whether the leftist parties "speak", as expected, more in terms of Benevolence, and the rightist more in terms of Tradition) or school textbooks (e.g. whether civic education materials stress the Self-direction values) when one is interested in the question of values embedded in the educational system. If or when a credible source of sufficient breadth and scope is

identified (e.g., a newspaper archive), the societal or population level values can be assessed as well, just in line with the Bardi et al. (2008) approach, only slightly covered in the present study.

Depending on the particular micro-/mezzo-/macro-focus, further possibilities for addressing the issues of the relationship between value preference and other relevant psychological phenomena (e.g. between values and behaviour, by measuring the presence of Benevolence and Universalism values in the essays describing reasons for prosocial behaviour, or Hedonism values in the description of leisure time activities practiced), group differences in values (e.g. promotion of values in different parties' manifestos, differences in the values embedded in different type of educational institutions – say, business versus medicine schools or faculties) or changes in societal values over time (e.g. differences in the prominence of Self-direction values before and after the so-called democratic changes in Serbia in 2000, or differences in the Tradition values promotion in history schoolbooks from the Socialist as compared to the pluralist period) become research questions that can be empirically addressed. Such a study would have some very important advantages over typical value research based on the self-report measures. Assessing values in archive and secondary materials – the products of social interaction made for some other, non-research "purposes" – implies studying real life phenomena right there where they happen, granting high ecological validity. Furthermore, there are additional possibilities of integrating the present lexicon of values into some other well-known text analysis procedures and software, such as the LIWC (Pennebaker et al., 2007), thus opening up new research opportunities (e.g., analysing whether the values measured in the text are related to the differences in frequency of other LIWC categories).

Limitations and recommendations for future research

There are several important limitations of the presented analyses which should be addressed in future research. Some of these stem from the informed and weighted, but, nevertheless, arbitrary decisions made during

the research process. The lexicon used in this study is but one of the many possible and, probably, equally valid sets of words in Serbian representing basic values. Hence, for example, whether to use the word '*uspeh*' or the word '*uspešnost*' – both equally good and acceptable translations of the English word '*success*' – was, in essence, an arbitrary decision. Bearing in mind that the lexicon of values used in this research is relatively short, it may not cover the full content of the basic values (see also Ponizovskiy et al., 2020), especially the refined theory of values proposed by Schwartz et al. (2012). Further research could benefit from using a larger set of words and lexicon refinements which would be more in line with the refined theory of values or include the main synonyms or words very similar to those that ended up in the final version of the value lexicon developed in this survey.

Additionally, Google searches, as well as the automated text analysis of this kind, remain 'blind' for all the language peculiarities which we had already mentioned earlier, such as the words' connotation, use of antonyms, differences in meanings across different contexts of language use etc. Language-related issues that are specific to Serbian warn further caution. The Serbian language has seven noun cases and two legitimately used scripts (Latin and Cyrillic). Although Google searches with, say, the noun[8] in the nominative case and the Latin script (as used in this study) report the number of hits that include both the Latin and Cyrillic and all the noun cases, possible differences in search results depending on the noun case and the used script deserve more attention in future studies[9]. Furthermore, Internet archives in

[8] Using nouns as value terms has additional theoretical significance because it serves to mainly capture the terminal values only, escaping all of the peculiarities with the differentiation between the terminal and instrumental values (Rokeach, 1973), which the most recent Schwartz's value model abandoned. Some of the mentioned value questionnaires used a small number of items to measure personal values and used nouns exclusively (e.g. SSVS, Lindeman & Verkasalo, 2005).

[9] During the study preparation and before conducting the Google searches reported here, we piloted different search scenarios, using, for example, the Cyrillic script or averaging the number of search results for the same noun in different cases. Although there were some differences in search results between these solutions and

Serbian (e.g. those of the leading media) cover a relatively short time span, and greatly vary in the amount of content and web infrastructure. Therefore, instead of solely relying on Google searches, future research would benefit from supplementing it with textual material and reliable corpora of different sources (such as the newspaper archive used in the original study or available language corpora in Serbian, e.g., srWAC and similar).

Similarly, further refinement and testing of different procedures and solutions for the estimations of value prominence should be probed in future research. For the present purposes, the performed analyses are sufficient to answer the research aims. But that would, in fact, be necessary if the value lexicon were to be used in the text materials that are organized by some other principle. For example, it could be adapted to some well-known concordance measures used in linguistic research which take into account the number of words in some limited text unit (e.g. normalized frequencies per thousand words or per million words). Some other networking and mapping technique could also be probed when analysing the patterns of co-occurrence in future research or for some other purposes.

Finally, while the lexicon development and lexical co-occurrence analysis followed the original procedure in full, and the analysis of its convergence with the self-report measure for the most part, the main limitation of the present study is related to the analysis of the lexicon's criterion validity. The analysis of the criterion validity of the value lexicon was only exploratory here and necessitates further studies, for which the present research, hopefully, offers a starting point and useful input. Along with the measures of values assessed by the value lexicon, using objective, and, preferably, behavioural measures obtained from independent sources when estimating criterion validity would be most welcome.

those that were finally implemented, these were almost negligent and suggested that there were no reasons not to adopt a more parsimonious solution. This is why, for example, we did not use a lemmatization strategy.

Funding

The work of the first author was financially supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, as part of the financial support of the scientific research at the University of Belgrade – Faculty of Philosophy (contract number 451-03-47/2023-01/ 200163).

Conflict of interest

We have no conflicts of interest to disclose.

Data availability statement

Data and materials needed for the reproduction of the results reported in the paper are available at: https://osf.io/nk8z7/?view_only=3c2327fb3db54f1c9b51e984cfdd15cb.

References

Araujo, T., Wonneberger, A., Neijens, P., & de Vreese, C. (2017). How much time do you spend online? Understanding and improving the accuracy of self-reported measures of internet use. *Communication Methods and Measures, 11*, 173–190. https://doi.org/10.1080/19312458.2017.1317337

Arieli, S., & Tenne-Gazit, O. (2017). Values and behavior in a work environment: taking a multi-level perspective. In S. Roccas & L. Sagiv (Eds.), *Values and Behavior*: *Taking a Cross Cultural Perspective* (pp. 115–141). Cham: Springer.

Bardi, A., Calogero, R. M., & Mullen, B. (2008). A new archival approach to the study of values and value – Behavior relations: Validation of the value lexicon. *Journal of Applied Psychology, 93*(3), 483–497. https://doi.org/10.1037/0021-9010.93.3.483

Blair, I. V., Urland, G. R., & Ma, J. E. (2002). Using Internet search engines to estimate word frequency. *Behavior Research Methods, Instruments, & Computers, 34,* 286–290. https://doi.org/10.3758/BF03195456

Chen, J., Hsieh, G., Mahmud, J. U., & Nichols, J. (2014). Understanding individuals' personal values from social media word use. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 405–414). ACM. http://doi.org/10.1145/2531602.2531608

Crompton, T. (2010). Common cause – The Case for Working with Our Cultural Values. Dostupno na: *https://assets.wwf.org.uk/downloads/common_cause_report.pdf.*

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. https://doi.org/10.1093/poq/nfn035

De Raad, B., Morales-Vives, F., Barelds, D. P., Van Oudenhoven, J. P., Renner, W., & Timmerman, M. E. (2016). Values in a cross-cultural triangle: A comparison of value taxonomies in the Netherlands, Austria, and Spain. *Journal of Cross-Cultural Psychology*, *47*(8), 1053–1075.

European Social Survey European Research Infrastructure (ESS ERIC). (2021). ESS9 - integrated file, edition 3.1 [Data set]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ESS9E03_1

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler, (Ed), *Review of Personality and Social Psychology* (pp. 141–165). Beverly Hills, CA: Sage.

Lazić, A., Lazarević L. B., Purić, D., & Žeželj, I. (2021). REPOPSI: The open repository of psychological instruments in Serbian. In N. Miljković, P. Pejović, & M. Cvetanović (Eds.), *Proceedings of the Third National Conference with International Participation titled Application of free software and open hardware PSSOH 2020* (pp. 51–56). University of Belgrade School of Electrical Engineering and Academic Mind. https://doi.org/10.5281/zenodo.4064293

Lindeman, M., & Verkasalo, M. (2005). Measuring values with the short Schwartz's value survey. *Journal of Personality Assessment*, *85*(2), 170–178. http://dx.doi.org/10.1207/s15327752jpa8502_09

Marušić-Jablanović, M. (2018). Value priorities of future preschool and class teachers in Serbia in terms of the Schwartz refined value theory. *Psihologija*, *51*(2), 143–161. https://doi.org/10.2298/PSI160902013M

Mukta, M.S.H., Ali, M.E., Mahmud, J. (2016). User Generated vs. Supported Contents: Which One Can Better Predict Basic Human Values?. In: Spiro, E., Ahn, YY. (eds), *Social Informatics* (pp. 454–470). Springer, Cham. https://doi.org/10.1007/978-3-319-47874-6_31

Pavlović, Z. (2021). *Psihologija vrednosti* [The Psychology of Human Values]. Beograd: Institut za psihologiju.

Pavlović, Z., & Stepanović Ilić, I. (May, 2021). Basic values and leisure-time activities among youth. Presented at the 27th scientific conference *Empirical Studies in Psychology*, Institute of Psychology, Laboratory for Experimental Psychology & Faculty of Philosophy, Book of Abstracts, pp. 115, Belgrade, 13–16th May.

Pavlović, Z., & Stepanović Ilić, I. (2022). Basic values as predictors of leisure-time activities among adolescents. *Primenjena psihologija*, *15*(1), 85-117. https://doi.org/10.19090/pp.v15i1.2349

Paulhus, D. L., & Vazire, S. (2007). The self-report method. *Handbook of research methods in personality psychology*, *1*(2007), 224-239.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77,* 1296–1312. https://doi.org/10.1037//0022-3514.77.6.1296

Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the lifespan. *Journal of Personality and Social Psychology, 85*(2), 291-301.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC2007*. Austin, TX: LIWC Inc.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54,* 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Ponizovskiy, V., Ardag, M., Grigoryan, L., Boyd, R., Dobewall, H., & Holtz, P. (2020). Development and validation of the personal values dictionary: A theory–driven tool for investigating references to basic human values in text. *European Journal of Personality*, *34*(5), 885-902. https://doi.org/10.1002/per.2294

Portman, A. (2014). *From might to mandate: Values in rulers' public speeches in Finland 1809-2000*. Helsinki University: Publications of the Department of Social Research. http://doi.org/10.2870/35887

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org.

Radović, O. (2010). *Socio-psihološki činioci vrednosnih orijentacija mladih srpska nacionalnosti na Kosovu i Metohiji – evaluacija teorije vrednosti Saloma Švarca*. Neobjavljeni magistarski rad. Beograd: Filozofski fakultet.

Radović, O., Stanojević, D., & Tošković, O. (2019). Values and subjective well-being in context of transition: Data from Serbia. *Sociološki pregled, 53*(3), 908–942. https://doi.org/10.5937/socpreg53-21692

Roccas, S., & Sagiv, L. (Eds.). (2017). *Values and behavior: Taking a cross cultural perspective*. Cham: Springer.

Roccas, S., Sagiv, L., & Navon, M. (2017). Methodological issues in studying personal values. In: S. Roccas & L. Sagiv (Eds.), *Values and behavior: Taking a cross cultural perspective* (pp. 15-50). Springer.

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

Sanderson, R., & McQuilkin, J. (2017). Many kinds of kindness: the relationship between values and prosocial behaviour. In S. Roccas & L. Sagiv (Eds.), *Values and Behavior*: *Taking a Cross Cultural Perspective* (pp. 75–96). Cham: Springer.

Saucier, G., & Goldberg, L. R. (1996). The language of personality: Lexical perspectives on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 21–50). Guilford Press.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 165). San Diego, CA: Academic Press. https://doi.org/10.1016/S0065-2601(08)60281-6

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues, 50*, 19–45. https://doi.org/10.1111/j.1540-4560.1994.tb01196.x

Schwartz, S. H. (2003). A Proposal for Measuring Value Orientations across Nations. Chapter 7 in the ESS Questionnaire Development Report.

www.europeansocialsurvey.org/docs/methodology/core_ess_questionnaire/ESS_core_questionnaire_human_

Schwartz, S. H., & Cieciuch, J. (2021). Measuring the Refined Theory of Individual Values in 49 Cultural Groups: Psychometrics of the Revised Portrait Value Questionnaire. *Assessment*. https://doi.org/10.1177/1073191121998760

Schwartz, S. H., & Ros, M. (1995). Values in the West: A theoretical and empirical challenge to the individualism– collectivism cultural dimension. *World Psychology, 1,* 91–122.

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J.-E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology, 103*(4), 663–688. https://doi.org/10.1037/a0029393

Sun, M., Zhang, H., Zhao, Y., & Shang, J. (2014). From post to values: Mining Schwartz values of individuals from social media. In Chinese National Conference on social media processing (pp. 206–219). Berlin, Heidelberg: Springer.

Temple Lang D (2022a). _RCurl: General Network (HTTP/FTP/...) Client Interface for R_. R package version 1.98-1.9, <https://CRAN.R-project.org/package=RCurl>.

Temple Lang D (2022a). _XML: Tools for Parsing and Generating XML Within R and S-Plus_. R package version 3.99-0.13, <https://CRAN.R-project.org/package=XML>.

Vecchione, M., Schwartz, S. H., Caprara, G. V., Schoen, H., Cieciuch, J., Silvester, J., Bain, P., Bianchi, G., Kirmanoglu, H., Baslevent, C., Mamali, C., Manzi, J., Pavlopoulos, V., Posnova, T., Torres, C., Verkasalo, M., Lönnqvist, J.-E., Vondráková, E., Welzel, C., & Alessandri, G. (2015). Personal values and political activism: A cross-national study. *British Journal of Psychology, 106*(1), 84–106. https://doi.org/10.1111/bjop.12067

Vesić Pavlović, T. (2013). Dominantne društvene vrednosti u javnom diskursu Srbije kroz prizmu kognitivne lingvistike [Dominant social values in the public discourse of Serbia through the prism of cognitive linguistics]. In: *Jezik, književnost, vrednosti: Jezička istraživanja* (pp. 327-337). Faculty of Philosophy, University of Niš.

Appendix A

A representative part of the raw co-occurrence matrix for the value words (the number of Google hits in 1,000)

	moć	snaga	kontrola	postignuće	ambicioznost	uspešnost	luksuz	zadovoljstvo	uživanje	uzbuđenje	raznovrsnost	stimulacija	nezavisnost
moć	0	5770	3100	331	315	1110	416	2050	1910	629	239	194	858
snaga	5770	0	5900	337	405	1540	763	2790	1790	748	297	384	2030
kontrola	3100	5900	0	197	240	678	382	1460	1110	338	196	249	753
postignuće	331	337	197	0	36.6	53.7	37.1	189	111	72.4	51.4	41.1	108
ambicioznost	315	405	240	36.6	0	146	58.5	232	130	68	81.9	52	153
uspešnost	1110	1540	678	53.7	146	0	130	481	341	121	50.1	77.5	257
luksuz	416	763	382	37.1	58.5	130	0	510	397	101	136	39.1	181
zadovoljstvo	2050	2790	1460	189	232	481	510	0	1560	523	395	300	516
uživanje	1910	1790	1110	111	130	341	397	1560	0	357	114	131	327
uzbuđenje	629	748	338	72.4	68	121	101	523	357	0	91.3	117	190
raznovrsnost	239	297	196	51.4	81.9	50.1	136	395	114	91.3	0	40.5	178
stimulacija	194	384	249	41.1	52	77.5	39.1	300	131	117	40.5	0	93.1
nezavisnost	858	2030	753	108	153	257	181	516	327	190	178	93.1	0
samostalnost	866	1520	832	89	127	237	128	459	371	111	83.8	103	585
sloboda	2180	2860	1670	211	285	413	441	1550	1170	386	337	185	2030
jedinstvo	965	1530	677	100	108	329	138	616	313	151	107	73.2	480
pravda	889	1820	679	73.7	114	180	142	512	320	117	118	74	667
jednakost	430	390	355	57.3	72.3	101	44.1	237	160	65.5	129	81.6	341
dobrota	526	695	325	53.8	72.5	88.4	59.1	395	224	112	99.9	56.7	352
velikodušnost	281	334	129	47.6	49.9	37.9	59.9	203	137	82.5	38.3	19.9	87.5
milost	752	947	395	53.4	106	131	102	502	237	146	109	66.9	440
tradicija	2110	2750	1320	145	233	456	310	1240	1440	302	449	165	763
običaji	1580	1930	787	64.6	102	391	250	654	666	166	205	92.6	571

poštovanje	2640	4200	2230	197	211	606	359	1930	1070	346	337	233	976
uzdržanost	403	591	358	46.7	33.9	91.8	58	238	169	39.7	81.5	44.7	292
uvažavanje	199	201	118	29.3	35.5	59.7	41.8	125	117	56.7	37.1	20.9	73.5
brižnost	132	194	130	9.44	44.6	51.7	28.4	107	51.4	30.3	41	44.8	115
bezbednost	1360	2740	1930	93.6	126	797	230	707	874	178	308	210	925
sigurnost	2310	6650	3260	209	240	339	310	1800	1580	406	236	222	914
zaštita	4960	11700	6310	263	271	1350	640	3010	2270	507	302	336	1310