IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 22, 2020, accepted February 23, 2020. Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2020.2976633

Construction of Human Motivational Profiles by Observation for Risk Analysis

ADAM SZEKERES AND EINAR ARTHUR SNEKKENES
Department of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology (NTNU), 2802 Gjøvik, Norway
Corresponding author: Adam Szekeres (adam.szekeres@ntnu.no)

This work was supported in part by the project IoTSec—Security in IoT for Smart Grids, part of the IKTPLUSS program funded by the Norwegian Research Council under Grant 248113/O70.

ABSTRACT This study aimed at analyzing the extent to which publicly observable pieces of information representing stakeholders' past and current choices can be utilized for the construction of motivational profiles. Motivation is operationalized by the theory of Basic Human Values, which organizes 10 values capturing distinct aspects of human motivation into a hierarchical order. The construction of motivational profiles for individual stakeholders is motivated by the need to enhance the existing decision-maker model in the Conflicting Incentives Risk Analysis (CIRA) method. This study utilized an online questionnaire to collect responses from participants (n = 331) about a wide range of habits and personal items that are easily observable in various contexts by an analyst. The validity of the set of observables as surrogate predictors of the motivational profiles is evaluated by various methods (i.e. comparison to previous results, cross-validation of models, comparison to test-retest reliability of the psychometric instrument) and techniques (calculation of prediction interval for individual profile scores). The assessment of the uncertainties associated with predicting motivational profiles is explored in detail. Additionally, an example illustrates how the profiles can be utilized for the assessment of action desirability (i.e. prediction of behavior) based on the utility calculations established in CIRA. The results contribute to an improved understanding about the accuracy with which human stakeholder motivation can be inferred from public observables and utilized within the context of information security risk analysis.

INDEX TERMS Conflicting Incentives Risk Analysis, human motivation, public observables, profiling, risk analysis, stakeholder behavior.

I. INTRODUCTION

Increasing levels of digitization affect more and more sectors, as well as critical infrastructures. A prominent emerging example is the Smart Grid, which represents the augmentation of the traditional electric grid with Internet of Things (IoT) devices enabling several desirable properties for various stakeholders, such as enhanced monitoring and control capabilities, the potential for more sustainable and eco-friendly energy utilization, new business opportunities, etc. [1]. The envisaged benefits can materialize given that the potential downsides introduced with novel technologies remain under control, and risks are mitigated. A complex system such as the Smart Grid has an inflated surface for cyber-attacks [2], and potential threats to privacy are increased [3]; thus, national security may be at risk when international conflict permeates to critical infrastructures [4].

Any stakeholder connected to the electric grid may be interested in getting answers to questions relating to their level of risk as a result of the actions or inactions of other parties. Homeowners may be interested in the privacy risks they face as owners of IoT smart appliances when service providers and manufacturers decide to merge [5]. Is it possible that threats to end-users' privacy observed in other sectors (e.g. toll booth use [6], health care [7]) are transferable to the Smart Grid when the need to process huge amounts of information at a low cost motivates companies to engage in outsourcing, which may expose millions of citizen records to parties whose interests may be difficult to monitor. Analysis of consumption data from Smart Meters enables profiling that can be used to identify unique devices

The associate editor coordinating the review of this manuscript and approving it for publication was Ana Lucila Sandoval Orozco.

used in the household, to reveal the number of occupants and other sensitive information not previously available from these sources of data. Such datasets are of great potential utility not only for the electricity provider, but other third parties (e.g. insurance companies, entertainment companies, and government authorities) [8]. Are the proper regulations in place, and are they enforced so that they prevent electricity companies from abusing their newly gained insights? Is it reasonable to assume that the information provided to prospective customers about the details of their contract is valid and reliable (i.e. integrity of information is appropriate)? Misinformation or deliberate withholding of pieces of key information can have a negative impact on the organization, when misbehavior is revealed [9]. What are the key factors that motivate relevant decision-makers in an organization to invest scarce resources (e.g. time and money) to ensure that customer information is securely transmitted, processed, stored and erased throughout the entire lifecycle of the data [10]? From a national security perspective, it is important to understand whether all the stakeholders responsible for maintaining and developing the Smart Grid of the future act in accordance with national interests.

In order to answer such questions in highly complex systems, the Conflicting Incentives Risk Analysis (CIRA) method proposes a way to model risks in a novel way by focusing on the motivation of the relevant stakeholders within the scope of the analysis [11]. The method requires that stakeholder motivational profiles be constructed without direct interaction with the subjects (i.e. using unobtrusive measures). Therefore, the main objective and **research problem** of this paper is to explore the extent to which publicly observable pieces of information can be utilized for building individual motivational profiles. For the construction of the motivational profiles, this study focuses on two distinct types of information that are assumed to be easily available in any context and can be assessed with a high accuracy simply by observation of subjects:

- evidence of conscious choices from the past of the stakeholder (i.e. ownership of various items, buying decisions) and
- habits and activities in the present.

There is a growing collection of work that demonstrates how various personality features can be predicted from different behavioral traces: intelligence and Big 5 traits from Facebook likes [12], Big 5 traits from mobile phone use data [13], and so on; for an extensive review, see [14]. While these methods are unobtrusive in the sense that they do not rely on direct interaction with the subject, they are highly obtrusive since they require access to sensitive personal account information or behavioral characteristics available only in settings where a subject explicitly gives permission to an application or other data collecting service, which can be used to amass a vast amount of information about the subjects. It is, however, unreasonable to assume that such sources of information will be available in common risk analysis settings. Furthermore, respondents would not be legally obliged to provide such account information for the purpose of the analysis. Therefore, the utility of the previously mentioned unobtrusive profiling methods is highly limited for the purpose of a real-world risk analysis, since the methods require full access to devices and/or accounts and are dependent on specific services. The present analysis focuses on features that are independent of service providers and thus aims for a wider range of applicability. The following **research questions** were formulated to address the overall research problem:

- **RQ 1:** How well can observable features predict stakeholder motivational profiles operationalized as the Basic Human Values?
- **RQ 2:** How much improvement can be expected from the present set of observable features compared to analyses using demographic features?

The paper is organized as follows. Section II presents a set of the most widely adopted information security risk assessment (ISRA) methods, which include human-related risks in the risk assessment procedure; describes key features of the CIRA method; and presents the theory used to operationalize human motivation. Section III describes the research method, including the data collection procedure and the instruments used. Section IV presents the key findings, Section V discusses the relevance of the findings in the context of risk analysis, as well as the limitations and plans for further work. Finally, Section VI provides a summary of the work.

II. RELATED WORK

Several risk analysis methods exist, but few address human motivation in detail. This section provides and overview of existing approaches for addressing human-related risks within information security as implemented in various ISRA methods. The primary resource for this overview is provided by [15], in which the most relevant ISRA methods are analyzed in detail, and the Core Unified Risk Framework is developed to aid practitioners in selecting the most appropriate method for the task at hand. The framework contains a total of nine ISRA methods, along with privacy and cloud risk assessment methods. The following overview focuses on a subset of the methods, including a discussion of human threats. Four methods were excluded from the present overview, due to their incompleteness on the following attributes according to the framework: threat willingness/motivation, threat capability, and threat capacity. Furthermore, the Risikovurdering av informasjonssystem (RAIS) method was also excluded due to its obsolescence and unavailability in English. A short summary of the strengths and weaknesses of the reviewed methods is provided in Table 1.

The **ISO 27005:2011** is one of the most widely used risk management frameworks, which in Annex C lists various threats that can guide an analyst through the threat assessment process [16]. Each *threat* is classified into one or more of the following groups: *accidental, deliberate*, and *environmental*.

TABLE 1. Comparison of a representative set of ISRA methods with respect to their capability of dealing with human threats.

No.	Method	Pros	Cons
1.	ISO 27005:2011	Wide acceptance in community. List of most relevant human threat groups. List of attributes associated with each identified group.	Lack of guidelines on how to assess the attributes, how to derive valid probabilities, consequences, etc.
2.	FAIR	Based on solid quantitative theories and methodology. The risk landscape development is supported by software tools.	Difficult to check that the obtained parameters are correct. Extensive training is required to apply the method.
3.	OCTAVE-Allegro	Suitable for preliminary assessments especially when resources for conducting a risk analysis are limited. Requires no expert knowledge. Variety of supporting tools.	Lack of quantitative rigour. No systematic way for threat discovery. No guidance on mitigation strategies against human threats.
4.	NIST 800-30	Threat characteristics: adversary intent; adversary capability; adversary targeting. Provides human threat sources in a taxonomy. Designed for the needs of federal information systems.	Unclear how assumptions about threat characteristics could be verified. Potential for generating unmanageable amounts of threat scenarios.
5.	COBIT5/RISK IT	Emphasis on aligning business objectives and risk analysis objectives through a focus on critical assets. Basic classification of threat types including malicious, accidental human threats.	Lack of instructions about the procedures for assessing human-related risks.
6.	CORAS	Risk analysis aided by graphical representations, focus on re-usability of previous results. Input from various stakeholders with different knowledge and experience.	Success of risk assessment largely depends on subjective evaluations, and on the experience of participants of brainstorming sessions. Elementary conceptualization of human threats.
7.	CIRA	Suitable for emerging systems (without historical data). Addresses Opportunity Risk. Redefines risk as the misalignment between stakeholder motivations.	Requires enhancement to enable real-world applicability by characterizing stakeholders. Lacks validation in real settings.

An additional table organizes human-related threats into five main groups by their origins (*hacker, computer criminal, terrorist, industrial espionage*, and *insiders*). Each group has an associated list of motivations, and the possible consequences of these threats are enumerated. Annex D also mentions several human-related vulnerabilities that span across issues related to personnel (e.g. lack of security awareness), organizational vulnerabilities (e.g lack of continuity plans), hardware and software (e.g. complicated user interfaces). The framework produces a risk matrix for further decision-making, which can be constructed by combining subjective and empirical measures, that fit well with the organization's objectives and available resources. In sum, the framework calls the analyst's attention to several human threats and provides general outlines about issues that should be considered during threat identification and vulnerability assessment, which can be useful during a high-level initial risk identification phase. However, the analyst is not provided with specific details about the complex motivational and cognitive processes that result in overt behavior. Since an analyst may have to resort to guesswork regarding human threats, the risk assessment procedure could result in ignoring or miscalculating human-related threats and risks.

The simulation-based Factor Analysis of Information Risk **(FAIR)** method was developed to measure and represent information security risk using quantitative methods and statistically sound mathematical calculations. Salient objects are identified in the environment, their characteristics are defined, and their interactions are modeled. The end result can be an integer, a distribution that represents the risk to information security. ''Information risk occurs at the intersection of two probabilities-the probability that an action will occur that has the potential to inflict harm on an asset, and the probable loss associated with the harmful event'' [17]. A taxonomy for information risk includes elemental components (objects) that make up the information risk landscape, a set of variables that describe the characteristics of objects, a decomposition of the factors that drive information risk, and a description of the relationships between the factors. Humans are a type of object within the framework, and *threat agents* are special objects that can be categorized as: *humans, animals, environmental elements, and human-made objects*. Threat agents, which have the ability or tendency to inflict harm upon other objects, are characterized by a unique set of characteristics that captures a certain level of psychological realism, including *skill, knowledge, experience, resources, risk tolerance, primary and secondary motives, and intents*. A *threat community* provides a description for a set of threat agents that share some common characteristics. It is useful for defining threat agent characteristics based on group membership when individuals are unknown. FAIR takes the perspective of the threat agent when considering the value of the object, the vulnerability of the object and the level of risk to the threat agent with negative consequences. These considerations are included based on the specific threat community characteristics. To measure *threat capability*, a scale is constructed by combining three factors: knowledge, experience and resources. In sum, the method models human agents as a group within a threat community, where the parameters related to the specific threat community are *volume, activity level, capability, risk tolerance, selectiveness, primary intent, and secondary intent*.

The Operationally Critical Threat, Asset, and Vulnerability Evaluation **(OCTAVE-Allegro)** method was designed to optimize information security risk assessments, considering limited resources for the task. The methodology guides

the analyst through the process by considering how people and technology contribute to business processes they support [18]. Several OCTAVE variants have been developed for the needs of organizations of various sizes. All variants aim at developing qualitative risk evaluation criteria, identifying assets that are crucial to the goals of the organization, identifying vulnerabilities and threats to those assets, and evaluating potential consequences to the organization if identified threats are realized. Some variants are based on workshops with interdisciplinary analyst teams or field experts. Allegro is specifically designed to guide risk assessment without extensive expert knowledge. The methodology is supported by worksheets and questionnaires. Human behavior is addressed in the *threat scenario identification* process, distinguishing between accidental and deliberate actions by human actors. Threats have the following properties: *asset, access, actor* (person who may violate security requirements), *motive* (intention of the actor), and *outcome*. Actors are further categorized according to their position as *inside, outside*. Threat identification largely depends on incidental background knowledge (e.g. ''John is the only employee who knows the production specs for producing widgets and he has been talking about leaving the company; if he does so, and the widget specs aren't obtained, we can't make widgets'' [18]), which is brought to the analyst's attention by the use of threat scenario questionnaires. The scenario-based qualitative threat identification can be useful to highlight important aspects where more investigation is needed, but largely depends on the analyst's creativity or motivation and available resources to distinguish between realistic and unlikely threat events.

The **NIST 800-30** method developed by the National Institute of Standards and Technology of the U.S. Department of Commerce [19] was designed to assist with conducting risk assessments for federal information systems and organizations, with the aim of providing senior executives the information needed to determine appropriate courses of action when considering the identified risks. The impact of human behavior is discussed in *threat sources* (characterized by intent and method targeted at the exploitation of a vulnerability), which enables the development of corresponding *threat scenarios*. Types of threat sources may be: hostile cyber or physical attacks; human errors of omission or commission; structural failures of organization-controlled resources; and natural and man-made disasters, accidents, and failures beyond the control of the organization. Furthermore, when discussing vulnerabilities in the broader context, various types of vulnerabilities are enumerated that can be linked to human behavior (e.g. lack of effective risk management strategies and adequate risk framing; poor intra-agency communications; inconsistent decisions; misalignment of enterprise architecture to support mission/business activities; external relationships, such as dependencies on particular energy sources, supply chains, information technologies, and telecommunications providers, etc.). The assessment of incident likelihood for adversarial threats is based on: adversary intent, adversary capability, adversary targeting. Table D-2 in the Appendix provides a detailed taxonomy of adversarial Threat Sources at various levels (individual, group, organization, and nation-state), with further distinctions, such as *outsider, insider, trusted insider*, etc. Tables D3 to D5 describe relevant adversarial features (i.e. *capability, intent, targeting*), with descriptions of the meanings of the accompanying qualitative values (very low to very high). A lack of further guidance on what evidence is needed to support the adversarial models could hinder the effective assessment of human-related threats, and may turn risk assessment into an ad-hoc exercise, without empirical evidence supporting the assumptions. The volume of Threat Scenarios that can be potentially generated from the checklists may further complicate the execution of successful risk assessments.

The **COBIT5/RISK IT** framework developed by ISACA describes a process model for the management of information technology-related risk [20]. The document emphasizes that risk management-related processes need to be connected to overall business objectives, and communication between stakeholders should be a continuous process. During risk-management activities the guidelines propose the development and use of risk scenarios to help overcome the challenges associated with identifying important and relevant risks amongst all that can possibly go wrong with the IT infrastructure. The scenarios are used during the risk analysis, where the frequency of a scenario actually happening and business impacts are estimated. Scenarios should contain *actors* (internal or external), *threat type* (malicious or accidental), *event* (e.g. disclosure of confidential information), *asset* (impacted by the event leading to business impact), and *timing*. The Risk IT framework is mainly focused on the management and governance perspective, and no further details are provided about how to assess human-related risks.

The graphical or model-based method for security risk analysis **CORAS**, was developed in order to provide a method that facilitates risk analysis by making previous results easily accessible and maintainable [21]. The method comprises a specific risk-modeling language, the step-by-step description of the risk-analysis process, and a software tool for documenting and maintaining the results of the analysis. It is based on meetings and workshops between relevant stakeholders and the analyst. Risks are identified through a process called structured brainstorming, where participants with different competences and backgrounds provide their perspectives on the target of analysis. The outputs of the brainstorming activities are threat diagrams where human threats (accidental or deliberate) are linked to vulnerabilities, threat scenarios and incidents. The next step focuses on risk estimation, in which participants provide likelihood estimates and consequence estimations for each threat scenario in the threat diagrams. For scenarios with difficult-to-estimate likelihoods, the analysis leader gives suggestions based on historical data, like security incident statistics or personal experience. Thus, risk assessment largely depends on subjective evaluations and on the breadth of knowledge

possessed by the stakeholders invited to the brain-storming sessions.

In order to reduce the complexities associated with conducting the aforementioned risk-analysis methods, and to overcome some of their limitations, the Conflicting Incentives Risk Analysis **(CIRA)** method proposes a novel way to describe the risk situation. CIRA develops a different concept of risk, where risk is the result of misaligned stakeholder incentives [22] and risk is analyzed from the perspective of an individual facing a risk. Two types of risks are distinguished: **threat risk** -undesirable events- and **opportunity risk** - desirable events not realized. To characterize these risks, CIRA requires the identification of two classes of stakeholders: the **risk owner** and the **strategy owner**. Each stakeholder is characterized by a set of utility factors, which capture important aspects of their overall utility (e.g. wealth, health, security, etc.). Actions available for the strategy owner can have a positive or negative impact on the risk owner's utility factors. Each action has a level of (un)desirability for the strategy owner, which (de)motivates the selection of that particular strategy (e.g. the prospect of a monetary reward). The analysis therefore requires a detailed description of the dependencies between stakeholders, their motivational profiles and the inference of perceived gains and losses from the perspective of the strategy owner to enable the assessment of action-desirability in order to completely characterize the risk situation. The method's real-world applicability is currently limited by the lack of procedures and guidelines for assessing stakeholder motivations in a reliable and valid manner. Consequently, the method requires extensive validation to gain acceptance in the professional community.

Based on the surveyed ISRA methods, it can be concluded that the impact of human behavior on the security of systems is widely recognized, as demonstrated by the inclusion of the issue in most well-established ISRA methods. However, a valid assessment and characterization of human-related risks is, to a great extent, missing in existing approaches. Moreover, most ISRA methods do not place human behavior at the center of the analysis. These issues may be attributed to the difficulties with operationalizing and measuring motivation, intention, capability and probability of goal execution (e.g. adversarial), and so on. While data may be abundant about potential internal threat actors through logging of various activities, the extent to which they are indicative of threat realization may remain unexplored. Additionally, most of the reviewed ISRA methods do not employ an interdisciplinary approach when investigating crucial aspects of human behavior when formulating risk estimations. This gap is partially addressed in this work by investigating how a specific motivational theory from psychology can be utilized to enhance the human model of the CIRA method, which exclusively focuses on deliberate (motivated) human behavior when addressing risks. Thus, the enhanced method could complement other methods with a more valid and practical characterization of human behavior supported by empirical data.

A. CAPTURING MOTIVATION: THEORY OF BASIC HUMAN VALUES

The theory of basic human values identifies 10 distinct values that serve as guiding principles throughout people's lives [23]. The values included in the theory are cross-culturally recognized, capture distinct aspects of human motivation, and each refers to desirable end goals that people strive for. The theory proposes a dynamic relationship among the values, which form a circular structure presented in Figure 1. When making a decision, values that are on opposite sides of the circle tend to conflict with each other, while adjacent values are more compatible, giving indications about the trade-offs a decision-maker may be willing to make. Thus, key decisions can be represented by combining an individual's value hierarchy with the expected outcome of an action (i.e. is the duty to increase shareholder value for a CEO, more important than the desire to act lawfully?). Individuals and groups can be meaningfully characterized by their value hierarchies. Values possess the following six core features: 1. ''Values are beliefs linked inextricably to affect. 2. Values refer to desirable goals that motivate action. 3. Values transcend specific actions and situations. 4. Values serve as standards or criteria. 5. Values are ordered by importance relative to one another. 6. The relative importance of multiple values guides action.'' [24]. Previous work has utilized the theory of Basic Human Values for operationalizing a Strategy Owner's motivation to enhance CIRA's real-world applicability [25].

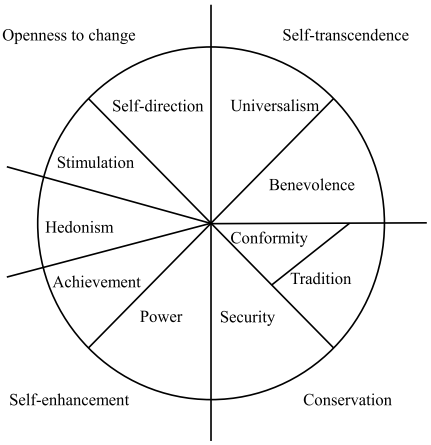

FIGURE 1. 10 basic human values with 4 higher dimensions forming a circular structure. Source: [24].

III. MATERIALS AND METHODS

This section provides a detailed description about the data collection procedures, the sample and the instruments utilized for gathering the necessary information from respondents to address the research questions of the study.

A. SAMPLE AND PROCEDURE

In order to reach a varied pool of respondents from the general population at a working age (above 18 years), a call

for participation was distributed on several online channels. As the main objective of the study was to assess the utility of publicly observable pieces of information for the construction of stakeholder motivation profiles, the following channels were used for participant recruitment to ensure a wide coverage of respondents with various backgrounds: A pilot study was conducted on Amazon Mechanical Turk to test the feasibility of data collection on the popular crowdsourcing platform; however it was assessed as inappropriate for the purpose of the present study since the majority of respondents are located in the U.S. (75%) and India (16%), working below median household incomes in the respective countries [26], which could hamper the transferability of conclusions to a different population (while not necessarily constraining model validation). Based on the results of the pilot study, modifications were implemented and links to the updated version of the survey were distributed on university and project-related mailing lists and on social media platforms. The survey was available in English and Norwegian, and the Norwegian translation was proof-read and finalized by a professional proofreading service. The data collection was open for a total of 114 days. The survey was implemented using the open-source Limesurvey tool and was hosted on internal servers provided by the university. The number of fully completed surveys is presented in Table 2, organized by the channels of recruitment.

TABLE 2. Number of completed surveys by distribution channels.

Distribution channel	Number of completed surveys
AmazonTurk	9
QR-invite	1
Social media	24
University e-mail list	332
Total	366

The validity of the final dataset was increased by removing responses below 10 minutes of completion time, which would indicate that respondents were not following the instructions carefully (estimated completion time: 20-30 minutes). Furthermore, extreme outliers with values exceeding three times the height of the boxes (25th-75th percentile) on each dependent variable's boxplot were identified and removed. The final sample **(n = 331)** consists of 173 males, 153 females, and 5 respondents with unspecified sex. The mean age is 40.28 years (SD = 13.27). Additional demographic descriptions of the sample are provided in Table 3. To compare the present sample to the general working-age population the information was obtained from the website of Statistisk Sentralbyrå (Statistics Norway) [27]. Compared to the general working-age population, in the present sample: males are slightly over-represented ($\approx 50\%$ in the population vs. $\approx 52\%$ in the sample), foreign citizens are over-represented ($\approx 11\%$ in the population vs. $\approx 25\%$ in the sample), the level of attained education is higher in the sample (tertiary: $\approx 37\%$ in the population vs. $\approx 68\%$ in the sample), and PhDs earned

TABLE 3. Basic demographic description of the sample.

Highest level of education completed	**n**	**Employment status**	**n**
Secondary school	26	Employed for wages	292
Bachelor's	53	Self-employed or homemaker	3
Master's	174	Student	23
PhD	78	Currently not in work	13
Citizenship		**Type of industry**	
Norwegian	247	Information and communication	26
Other	84	Professional, scientific and technical activities	146
Marital status		Public administration and defence; compulsory social security	33
Single	79	Education	78
Married or in a long-term relationship	235	Human health and social work activities	14
Divorced or separated	17	Other	34

was higher ($\approx 1\%$ in the population vs $\approx 23\%$ in the sample). The ratio of employed/unemployed respondents is similar to the ratio found in the population considering the active workforce ($\approx$ 3.7% in the population vs. $\approx$ 3.9% in the sample).

B. MEASURES

1) MOTIVATIONAL PROFILE - PVQ-21

Motivational hierarchy was assessed using the 21-item Portrait Value Questionnaire (PVQ). The PVQ was designed to measure the 10 basic value orientations and presents respondents with concrete and cognitively less-demanding tasks than previous instruments designed for measuring value structures. This makes the scale suitable for all segments of the population [28]. The PVQ includes short verbal descriptions of people with their goals and aspirations without explicitly identifying the values under investigation. Respondents answer by judging their own similarity to the portraits, and similarity judgments are transformed into a six-point numerical scale (reverse coded from the original as follows: 1 *(not like me at all)* to 6 *(very much like me)*). The PVQ's adequacy for measuring value structures is supported by adequate psychometric properties based on studies in several countries, and it is suitable for various forms of administration (e.g. face-to-face, by telephone, and online). As individuals may differ in their use of the response scale, centered scores were computed to correct for individual differences in response scale use, thus reflecting the relative importance of each value in the value system [29]. The original English version and the Norwegian version from the European Social Survey was used in this survey [30].

2) EVERYDAY CHOICES AND HABITS QUESTIONNAIRE

The next section of the survey collected information about various items and habits that are publicly observable. The aim of this part of the survey was to cover a wide range of items that can be observed in any situation without interaction with a respondent. Assessment of item ownership requires a single observation, while the assessment of habits may require observation over a longer period. The list of categories and the number of attributes collected per category

are presented in Table 4. Questions designed to assess habits asked respondents to report the approximate frequency of the activity for the last year. Other questions used yes/no questions, numerical input, or a single-choice format. The PDF version of the survey (in English) is available as supplementary material. Note that some differences between the original online version of the survey and the PDF version may exist as a result of exporting and converting it into a different format.

TABLE 4. Categories of publicly observable pieces of information collected from respondents, with number of attributes per category.

Ownership		Habits	
Home	4	ConsumptionPreferences	17
MeansOfTransport	23	FreeTimeActivities	26
ITdevices	21	Style*	5
Accessories	14	DietChoice*	1
Pets	6	SportsActivities	17
Tattoo	8	MusicPreferences	14
SocialMediaPresence	11	ClothingChoices	23
Jewelery	11	**BasicDemographics**	8
SportEquipments	16		

* Corresponding questions were not formulated to assess frequency of activity.

IV. RESULTS

The dependent variables (DV) of interest are the 10 Basic Human Values, ground truth scores collected by the PVQ-21 instrument. Data on a total of 225 independent variables were collected, which resulted in 437 variables after categorical (i.e. nominal) variables were recoded into indicator variables (where 0 = no/attribute is not present for the respondent; 1 = yes/attribute is present). This procedure is recommended so that categorical variables with several levels can be included in regression models. Reliability of the instrument was tested through the internal consistency measure (Cronbach's alpha), by analyzing all items that measure the same value. Cronbach's alpha measures the extent to which certain items of a test measure the same construct by analyzing the inter-relatedness of the items [31]. The analyses provided the following Cronbach's alpha scores for the 10 values: Conformity: .60, Tradition: .67, Benevolence: .56, Universalism: .52, Self-Direction: .36, Stimulation: .72, Hedonism .69, Achievement: .74, Power: .36, Security: .47. These results are similar to the reliability scores found in various nations [28]. It should be noted that the low alpha scores obtained in this and other studies (using the same instrument) may be attributed to the small number of questions (two or three for each dimension) measuring the same construct, which can decrease alpha scores [31]. By convention, alpha scores above .70 are preferred; however, there are no gold-standard levels of alpha, so even lower scores (.50) may be useful [32]. All the analyses were conducted using SPSS 25 by IBM, and scikit-learn, which is a free machine learning library for Python.

A. FEATURE SELECTION AND COMPARISON WITH PREVIOUS RESULTS

Following data pre-processing for each DV (10), several multiple linear regression models were built using the stepwise feature selection method in SPSS. This method searches among all independent variables that are not yet in the equation for the one which has the smallest probability of F ("The F-value is equivalent to the square root of the Student's t-value, expressing how different two data samples are, where one sample includes the variable and the other sample does not" [33]), and enters them into the equation if the inclusion criterion is met (p of entry set to < 0.05). Predictors in the regression equation were removed when their probability of F reached the criterion of exclusion (p of exclusion set to ≥ 0.1). The method stops when no predictor meets the inclusion/exclusion criteria. By tuning the exclusion and inclusion criteria, it is possible to control the final model's complexity. The procedure resulted in several models with increasing numbers of predictors and increasing levels of goodness of fit (R^2) associated with each model. The final set of predictors to be evaluated in the following step was selected from the model with the highest R^2 metric for each DV. R^2, or the coefficient of determination is calculated as $R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$, where SS_{res} is the sum of the residual squares and SS_{tot} is the total sum of squares, ranging between negative infinity and $+1$, which is a measure to assess the model's goodness of fit [34]. Table 5 summarizes details of the multiple linear regression models for each dependent variable. Adjusted R^2 scores represent a modified version of the R^2, which increases only when the additional terms improve the model more than expected by chance. Due to penalizing additional predictors the adjusted R^2 scores are always lower than corresponding R^2 scores for the same model. F-scores represent each model's improvement compared to the intercept-only models; df (degrees of freedom) signifies the number of predictors in each model.

TABLE 5. Summary of multiple linear regression models for each dependent variable.

	R^2	Adjusted R^2	F	df
Conformity	0.33**	0.29	311	19
Tradition	0.44**	0.39	303	27
Benevolence	0.36**	0.32	308	22
Universalism	0.40**	0.35	306	24
Self-Direction	0.60**	0.54	282	48
Stimulation	0.40**	0.36	309	21
Hedonism	0.49**	0.44	299	31
Achievement	0.48**	0.42	295	35
Power	0.47**	0.42	300	30
Security	0.40**	0.35	303	27
$^{**}p < 0.01$				

Figure 2 reports the performance of each model (red bars). Grey bars represent a the predictive utility of demographic features for building motivational profiles established

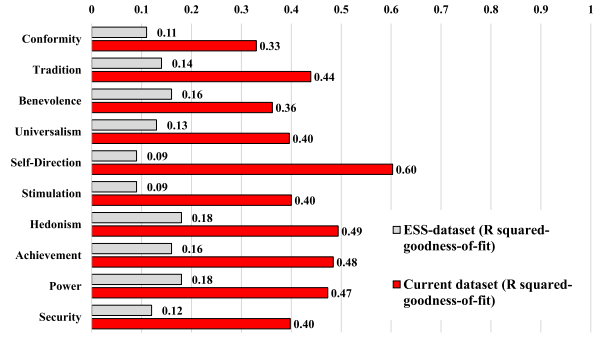

FIGURE 2. Prediction accuracy of Basic Human Values in terms of the R^2 metric. Red bars represent the maximum accuracy achieved after the models were built with the Stepwise feature selection algorithm in SPSS. Grey bars show the goodness of fit metrics for the same variables using demographic features from [25].

on the European Social Survey (ESS) dataset [25]. Differences between red and grey bars indicate the improvement between reported metrics from the two datasets. Across all DVs an average of 3.4-fold improvement is achieved by using the present class of predictor variables based on the R^2 metrics. Improvement for each DV was calculated as: $(R^2_{current}/R^2_{ESS})$, with AverageImprovement = Sum of improvements for each DV/10.

B. MODEL VALIDATION: EXPECTED PERFORMANCE ON UNSEEN DATA

The train-split re-sampling method was used to assess the proposed predictors' usefulness for predicting unobserved data points. The next set of experiments aimed at establishing the reliability of the regression models. This enabled the assessment of the model's performance on unseen data. Common practice is to evaluate the model, using only the goodness-of-fit metric; however, this generally leads to over-fitting, and cross-validation is rarely conducted in social science research [35]. ''Stepwise regression and all subset regression are in-sample methods to assess and tune models. This means the model selection is possibly subject to overfitting and may not perform as well when applied to new data.'' [36]. In order to assess the model's predictive performance on previously unseen data, various validation techniques can be used. Due to the small sample size, validation was achieved by conducting several train-test split validations, which is a form of validation with replacement, where the model is trained on a random 80% partition of the dataset, and the predictive performance is evaluated on the remaining 20% that was not utilized for model training. This procedure was repeated 100 times to assess the overall performance more accurately. Figure 3 reports the goodness of fit metrics for each dependent variable in terms of R^2 scores. Since the predictions are not made on the part of the dataset which was used for training the model, a decrease in predictive accuracy is to be expected, which is represented by the difference between the grey (i.e. models without validation) and red bars (i.e. models with cross-validations). Table 6 provides the list of the top five predictors for each dependent variable.

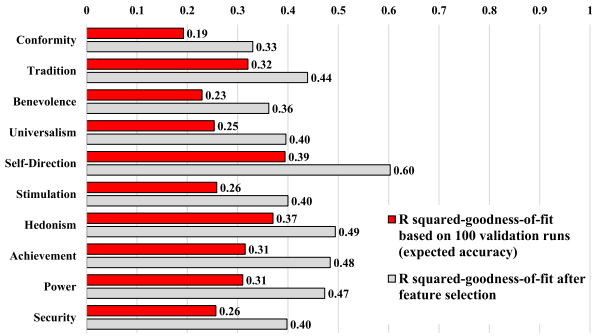

FIGURE 3. Prediction accuracy of Basic Human Values in terms of the R^2 metric. Grey bars represent the maximum accuracy achieved after the models were built with the Stepwise feature selection algorithm in SPSS. Red bars indicate the expected (mean) accuracy of the models after validation using 100 train-test split iterations.

C. MODEL PERFORMANCE EVALUATION AGAINST PVQ-21 TEST-RETEST RELIABILITY

Following a similar approach which was utilized in [12] for assessing prediction accuracy, Figure 4 compares the accuracy of predicting the Basic Human Values scores expressed by the Pearson product-moment correlation coefficients between ground-truth and predicted scores (red bars), whereas grey bars represent the accuracy of the PVQ-21, when the same test is taken by the same individuals (test-retest reliability), and the resulting scores are correlated with each other. The reference PVQ-21 reliability scores are derived from [28], in which a German student sample completed the PVQ-21 two times, separated by an interval of

TABLE 6. Top five features for predicting each dependent variable. Standardized Beta coefficients represent each independent variable's effect on the DVs.

Conformity		**Tradition**		**Benevolence**		**Universalism**		**SelfDirection**	
Activity_cigar	-0.18	Activity_presentation	-0.33	Item_carBrand	0.34	Activity_political	0.21	Activity_party	-0.26
Activity_music_alternative	-0.17	Item_numberOfCars	0.27	Activity_music_soundtrack	0.28	Activity_music_jazz	0.19	Item_carType_a	-0.25
Item_headphoneBrand	-0.17	Item_iceSkate	0.24	Activity_charity	0.21	Item _homeLocation	0.17	Item_carType_b	-0.24
Activity_music_electronic	0.15	Activity_highHeels	-0.21	Item_ski	-0.19	Activity_onlinePublishing	-0.17	Item_browser	-0.21
Activity_hunting	0.15	Item_socialMedia	-0.19	Item_searchEngine	0.17	Item_bicycleBrand	0.16	Item_headphoneBrand	-0.21
Stimulation		**Hedonism**		**Achievement**		**Power**		**Security**	
Activity_interview	0.27	Item_bicycleType	0.23	Activity_earring	-0.23	Demographic_citizenship	0.28	Activity_coffee	-0.27
Activity_cigarette	0.24	Item_homeOwnership	0.23	Activity_music_folk	-0.20	Activity_jacket	-0.25	Item_homeOwnership	0.21
Activity_music_alternative	0.18	Activity_fishing	-0.20	Item_bicycleType	-0.19	Item_phoneType	0.22	Activity_interview	-0.19
Activity_onlineForum	-0.18	Activity_learning	-0.19	Item_motorChoice	0.18	Item_homeLocation	-0.20	Activity_music_heavymetal	-0.19
Item_surf	0.15	Activity_snus	0.17	Item_laptopOS	0.17	Item_phoneColor	0.18	Activity_waterpolo	0.17

six weeks, to assess the reliability of the questionnaire. The test-retest reliabilities obtained in that study were moderate to high.

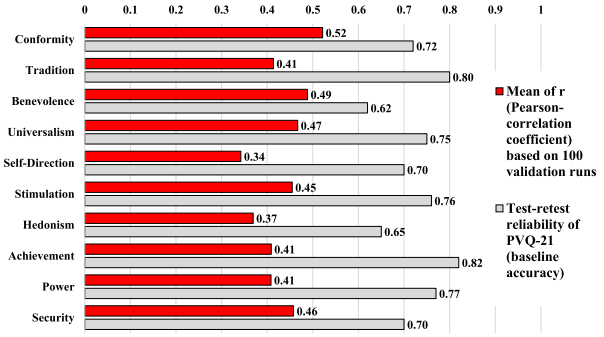

FIGURE 4. Prediction accuracy of Basic Human Values in terms of the Pearson correlation coefficients between predicted and ground-truth scores (red bars). The test-retest reliability is measure of correlation between the results of the PVQ-21 taken at different times by the same respondents (grey bars).

In the present sample Conformity achieved the highest accuracy (r = 0.52), followed by Benevolence (r = 0.49), Universalism (r = 0.47), Security (r = 0.46), Stimulation (r = 0.45), Power, Achievement, Tradition (r = 0.41), Hedonism (r = 0.37), Self-direction (r = 0.34) expressed in terms of the Pearson correlation coefficient between ground-truth and predicted scores. The absolute difference is smallest for Benevolence and Conformity; thus, these models can predict the related concepts nearly as well as the PVQ-21 questionnaire, while for the other values, each regression model achieves around half the accuracy of the original questionnaire. Table 7 complements Figure 4 by providing the mean goodness-of-fit and model-accuracy metrics for each dependent variable. In addition, one-sample Kolmogorov-Smirnov (K-S) tests were run on all metrics to assess whether the distribution of metric scores follows a normal distribution. Cases that do not follow a normal distribution are marked with *.

TABLE 7. Measure of goodness of fit (R^2) and measure of prediction accuracy (r - Pearson-correlation coefficient between ground truth and predicted scores) over 100 train-test split iterations.

	Mean of R^2 measures	SD	Mean of r Pearson-correlation coefficients	SD
Conformity	0.192*	0.103*	0.522	0.086
Tradition	0.320	0.115	0.414	0.083
Benevolence	0.229	0.104	0.488	0.079
Universalism	0.253	0.100	0.467	0.072
Self-Direction	0.124	0.124	0.342	0.077
Stimulation	0.259	0.104	0.455*	0.080*
Hedonism	0.370	0.105	0.369	0.069
Achievement	0.315	0.112	0.409	0.072
Power	0.310*	0.123*	0.408*	0.076*
Security	0.257*	0.095*	0.458	0.072

* denotes cases where normality hypothesis was rejected by the one-sample Kolmogorov-Smirnov test.

The corresponding K-S test scores are as follows: R^2 scores for Conformity D(100) = 0.099, p = 0.016, Power D(100) = 0.124, p = 0.001, Security D(100) = 0.105, p = 0.008; r-scores for Stimulation D(100) = 0.097, p = 0.02 and Power = D(100) = 0.096, p = 0.022.

D. EXAMPLE OF PREDICTING A SINGLE INDIVIDUAL'S PROFILE SCORES

Based on the formula for multiple linear regression:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_k X_k + \hat{\epsilon} \quad (1)$$

Table 8 shows a working example of how an individual's Conformity score is predicted using the trained model. The codes and associated meaning for frequency (i.e. habits) and dummy variables (i.e. ownership of item/existence of attribute) are summarized in Table 9.

TABLE 8. Prediction of an individual's Conformity value based on 19 features. PI - Prediction Interval refers to the individual prediction error.

Predictors (valid scores)	**Unstandardized Coefficient** β	**RawScore** (X_k)
Constant	0.218	
consumptionCigar (0-8)	-0.342	0
musicPreferenceAlternative (0-8)	-0.054	7
headPhoneBrand X (0-1)	-0.908	0
musicPreferenceElectronic (0-8)	0.050	1
sportsActivityHunting (0-8)	0.166	0
boatOwned (0-1)	-0.394	1
watchOwned (0-1)	0.291	1
clothWearSuit (0-8)	-0.073	2
laptopBrowser X (0-1)	0.545	1
bicycleBrand X (0-1)	-0.251	0
musicPreferenceSoundtrack (0-8)	-0.044	4
petsSmallMammal (0-1)	-1.428	1
socMedia X (0-1)	-0.222	1
tattooFig X (0-1)	-0.481	0
activityFrequency hairdresser (0-8)	0.073	5
phoneCoverColor X (0-1)	-0.395	1
carEnergy X (0-1)	-0.492	0
carType X (0-1)	0.233	1
watchBrand X (0-1)	0.419	1
IndividualPredictionError-Mean	0.077 (SD: 0.794)	
$\mathbf{Y}_{predicted}$ (95% CI)	**(-1.097) - (-0.785)**	

TABLE 9. Explanation of raw variable scores.

Code	**Meaning**
0	Never in the last 12 months
1	Once in the last 12 months
2	Twice in the last 12 months
3	Three to six times in the last 12 months
4	Seven to 11 times in the last 12 months
5	One to three times a month
6	Once or twice a week
7	Three or four times a week
8	Every day or nearly every day
0	No
1	Yes

A prediction interval (PI) captures the uncertainty around the predicted score, which is attributed to uncertainty of coefficients and additional error of individual data points. The errors of individual point estimates are calculated using the residuals from the predicted values using the bootstrapping sampling method (number of re-sampling = number of observations). A bootstrap sample was taken from the data, the model was trained, and a new outcome was predicted. A random residual was taken from the original regression fit and added to the new value. The procedure was

repeated for 100 iterations, and the resulting distribution of error terms was used to construct a variable with normal distribution that can be sampled randomly to capture the necessary error terms inherent in individual predictions (ϵ, PI) [36]. The one-sample Kolmogorov-Smirnov test did not reject the null-hypothesis (i.e. PIs are normally distributed $D(100) = 0.081$, $p = 0.101$).

E. EXAMPLE SCENARIO TO ASSESS ACTION DESIRABILITY

This section provides a simple example to assess the desirability of an action, which demonstrates how the method makes predictions about potential choices based on the derived value structure.

Predicted scores must be normalized by summing across all dimensions, then each score needs to be divided by the sum of scores, to quantify each value's relative importance. Formula: $w_i' = \frac{w_i}{\sum_{j=1}^{n} w_j}$ Thus, the normalized profile scores provide the necessary weights in Table 10. For the purpose of demonstration, the relative importance of values is taken from the pan-cultural empirical norms presented in Table 6 in [37]. The Strategy Owner faces a dilemma whether to implement an unconventional strategy that would provide significant personal gains and recognition from the organization's leaders, but which entails a misuse (secondary use) of customer data. An example of such a strategy considered by a stakeholder at an electric distribution system operator would be to use the detailed electricity consumption profiles of homeowners to infer their home occupancy patterns for promoting a novel home-surveillance service through personalized advertisements. Thus, the dilemma can be represented as *Option 0:* Do nothing - contributes positively to Conformity (i.e. restraint of actions that would harm or upset others), whereas Achievement values are unaffected; or *Option 1:* Implement strategy - contributes negatively to Conformity and contributes positively to Achievement (i.e. striving for personal success and recognition) values. For simplicity, the other utility factors are assumed to be unaffected by the choice. In order to compute the desirability of each option for the Strategy Owner, the Multi-Attribute Utility Theory is used as proposed in [22], where the overall utility of an option is calculated as the weighted sum of the individual utility factors using the formula: $U = \sum_{k=1}^{m} w_k \cdot u(a_k)$, in which m equals the number of utility factors of the stakeholder; w_k is the derived weight of utility factor a_k while $U = \sum_{k=1}^{m} = 1$; and $u(a_k)$ is the utility function for the utility factor a_k. Thus, to compute the utility of an option the normalized weight of each utility factor is multiplied by the score that represents the contribution of that choice on that particular utility factor (i.e. Initial Value, Option 0 - Final Value, Option 1 - Final Value) and these products are summed over all utility factors. For demonstration, the utility calculation for **Option 1** is as follows: $0.11 \cdot 40 + 0.07 \cdot 50 + 0.12 \cdot 50 + 0.11 \cdot 50 + 0.12 \cdot 50 + 0.09 \cdot 50 + 0.1 \cdot 50 + 0.11 \cdot 90 + 0.06 \cdot 50 + 0.11 \cdot 50 = 53.20$. The process is repeated for each identified decision option to enable the comparison between the desirability of various actions. Table 11 presents the overall utility calculations for the identified options. The Strategy Owner is assumed to be utility maximizing, therefore selecting the option with the highest overall utility (Option 1). The differences between the utilities associated with the Initial State, Option 0, and Option 1 can be interpreted as the strengths of motivation at work when the Strategy Owner contemplates a particular course of action.

TABLE 10. Expected effects of implementing a strategy on the relevant utility factors. Affected utility factors are marked in bold.

Strategy Owner's options:			**Option 0**	**Option 1**
Utility Factors	Normalized Weights	Initial Value	Final Value	Final Value
Conformity (%)	0.11	50	**70**	**40**
Tradition (%)	0.07	50	50	50
Benevolence (%)	0.12	50	50	50
Universalism (%)	0.11	50	50	50
Self-Direction (%)	0.12	50	50	50
Stimulation (%)	0.09	50	50	50
Hedonism (%)	0.10	50	50	50
Achievement (%)	0.11	50	50	**90**
Power (%)	0.06	50	50	50
Security (%)	0.11	50	50	50

TABLE 11. Overall utilities associated with the initial state and with making a choice. The outcome with the greatest expected utility is assumed to be selected by the Strategy Owner (i.e. Option 1 in this example).

Overall utility in Initial State	50.00
Overall utility of Option 0	52.11
Overall utility of Option 1	**53.20**

V. DISCUSSION

Modern societies keep on designing and implementing complex systems to fulfill certain goals with increasing efficiency (e.g. legal systems, markets for trading, voting, etc.). Most systems critical for modern life are enabled and dependent on innovations from information and communication technologies. The field has developed a variety of risk assessment methods and tools to deal with unexpected events by assessing the probability of such events and the consequences associated with them. Relatively less attention has been given to the consciously active part of the system - the human decision-maker with its unique motivations. This work aimed at improving the state of knowledge in relation to modeling human decision-makers for the purpose of risk analysis. More specifically, the study aimed at exploring the usefulness of easily observable pieces of information connected to potential decision-makers for inferring individual motivational profiles. This aim is supported by the requirements of the Conflicting Incentives Risk Analysis (CIRA) method, which uses stakeholder motivation to characterize risks. The results present the extent to which these features are valid predictors of the motivational profiles operationalized as the Basic Human Values. Furthermore, the results showed the added utility of this set of features in comparison to previous results using demographic data for the same purpose [25]. The reliability of profile predictions was assessed by various techniques (i.e. cross-validation, comparison with the

personality test's test-retest reliability, and calculation of prediction error (prediction interval) for predicting an individual's score). Some aspects of the motivational profile can be predicted nearly as well from the observable features as from the original psychometric instrument (Conformity and Benevolence).

While various steps were taken to include a diverse sample within the data collection, the relatively small sample size can be considered an important limitation, when the generalizability of the findings is considered. In replication studies, it would be desirable to have at least 10-20 unique observations for each category of the independent variables to ensure that inferences made from the sample are valid and robust for the target population. The external validity of the results could be improved using strict probability sampling, since most of the respondents were recruited through the university's e-mail list, which may result in a biased sample. Furthermore, the length of the survey needs to be reduced to increase respondent retention. Future studies may benefit from converting the obtained categorical data (e.g. type of phone) into corresponding retail prices to enhance the information content of the independent variables. The suitability of the established method for capturing action desirability for the stakeholders (i.e. computing the utilities according to Multi-Attribute Utility Theory) has to be investigated in future work. Choices of human stakeholders can be analyzed in real-world or in experimental settings to assess the procedure's applicability for capturing stakeholder intentions in various choice situations. The procedure's correctness would be verified if the investigation reveals a high degree of overlap between predicted (calculated on the basis of utility calculations) and actual choices made by subjects.

The agenda proposed in [35] calls for a shift in research strategy for psychology, with an increased focus on the prediction of behavior as opposed to explanation. The paradoxical state in which a good explanatory model is not necessarily good at predicting real-world behavior needs to be considered. While the objectives of the traditions may be different, methodological issues are enumerated as the reason for the discrepancy (e.g. p-hacking or lack of model validation on out-of-sample data). The paper proposes that the methodological shift should be aided by relying on machine learning (ML) methods that have been designed and used efficiently in various fields of computer science for the explicit purpose of generating predictive models that perform well on unobserved data as well. It is important to note that the present study utilized a traditional data analysis technique (using cross-validation to ensure reliability) instead of a complex ML method. This represents a conscious choice, where the transparency and interpretability of a simpler model is given a higher priority than the potential predictive improvement enabled by a complex ML method, which operates as a black box. The potential dangers of using black-box models for predictions affecting humans may result in gender or racial bias in case of school admission decisions [38], decisions about risk of re-offending behavior, risk of illness estimations, etc. [39]. Furthermore, European legislation also requires that algorithmic decisions that "significantly affect" subjects are to be explainable [40]. Easy interpretation of the model may increase the risk of manipulation and deception by motivated subjects, which has to be considered for real-world applications [41].

VI. CONCLUSION

This paper aimed at investigating the relevance of observable personal items and habits (public observables) for the construction of stakeholder motivational profiles. The stakeholder profiling method presented in this work is expected to complement the CIRA method, which focuses on stakeholder motivation to characterize risks within the domains of privacy and information security. The real-world applicability of the method depends on the accuracy with which stakeholder motivational profiles can be constructed without direct access to subjects. This paper assessed the predictive accuracy of publicly observable pieces of information associated with individual choices. It was demonstrated that these features are significantly better for profile-building than the most basic features that can be assessed by observation in any context (i.e. demographic features). Several comparisons and evaluations have been presented to assess the validity and reliability of the resulting profiles, and the uncertainty associated with the resulting profile scores has been assessed by the bootstrapping method (i.e. calculation of Prediction Intervals). The error associated with each predicted motivational score is modeled as a random variable with corresponding parameters from a normal distribution. Finally, a demonstration was presented using the utility calculations proposed in CIRA to assess the desirability of the options as perceived by the Strategy Owner in a potential choice situation. The presented work's main contribution is an enhanced understanding of the applicability of stakeholder motivational profiling for the purpose of risk analysis.

ACKNOWLEDGMENT

A. Szekeres would like to thank Dóra Szekeres for the initial Norwegian translation of the survey, Vebjørn Slyngstadli and Eigil Obrestad, for their help with setting up the hosting service and Jag Mohan Singh for useful discussions. The authors would like to thank the reviewers for their valuable comments which improved the overall quality of the paper.

REFERENCES

[1] O. B. Fosso, M. Molinas, K. Sand, and G. H. Coldevin, "Moving towards the smart grid: The norwegian case," in *Proc. Int. Power Electron. Conf. (IPEC)*, May 2014, pp. 1861–1867.

[2] A. Hahn and M. Govindarasu, "Cyber attack exposure evaluation framework for the smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 835–843, Dec. 2011.

[3] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Secur. Privacy Mag.*, vol. 7, no. 3, pp. 75–77, May 2009.

[4] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 Ukraine blackout: Implications for false data injection attacks," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3317–3318, Jul. 2017.

[5] K. Wiggers. 2018. *iRobot Partners With Google to Improve Smart Home Devices With Indoor Maps*. Accessed: Nov. 11, 2019. [Online]. Available: https://venturebeat.com/2018/10/31/irobot-partners-with-google-to-impro%ve-smart-home-devices-with-indoor-maps/
[6] M. K. Vigsnæs, P. Svaar, and V. Venli. (2019). *Slike Bilder Sender Bomselskap Til Kina: Nå går Datatilsynet inn Isaken*. Accessed: Nov. 7, 2019. https://www.nrk.no/norge/slike-bilder-sender-bomselskap-tilkina_-na-gar-datatilsynet-inn-i-saken-1.14754918
[7] A. C. Remen and L. Tomter.(2017). *Helse sør-Øst: Innrømmer at Utenlandske It-Arbeidere Fikk Tilgang Til Sensitive Pasientdata*. Accessed: Nov. 7, 2019. [Online]. Available: https://www.nrk.no/norge/helse-sor-ost_-innrommer-at-utenlandske-itarbeidere-har-hatt-tilgang-til-pasientjournaler-1.13478443
[8] R. R. Mohassel, A. Fung, F. Mohammadi, and K. Raahemifar, "A survey on advanced metering infrastructure," *Int. J. Elect. Power Energy Syst.*, vol. 63, pp. 473–484, Dec. 2014.
[9] J. Nordstrøm. (2017). *Forbrukerombudet:–Strømselskap Driver Med Ulovlig Telefonsalg*. Accessed: Nov. 7, 2019. [Online]. Available: https://e24.no/privatoekonomi/i/naMbeL/forbrukerombudets troemselskap-driver-med-ulovlig-telefonsalg
[10] B. Krumay, "The e-waste-privacy challenge," in *Privacy Technologies and Policy* (Lecture Notes in Computer Science), vol. 9857, S. Schiffner, J. Serna, D. Ikonomou, and K. Rannenberg, Eds. Cham, Switzerland: Springer, 2016.
[11] L. Rajbhandari, "Risk analysis using 'conflicting incentives' as an alternative notion of risk," Ph.D. dissertation, Gjøvik Univ. College, Gjøvik, Norway, 2013.
[12] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 15, pp. 5802–5805, Mar. 2013.
[13] N. Gao, W. Shao, and F. D. Salim, "Predicting personality traits from physical activity intensity," *Computer*, vol. 52, no. 7, pp. 47–56, Jul. 2019.
[14] M. Settanni, D. Azucar, and D. Marengo, "Predicting individual characteristics from digital traces on social media: A meta-analysis," *Cyberpsychology, Behav., Social Netw.*, vol. 21, no. 4, pp. 217–228, Apr. 2018.
[15] G. Wangen, C. Hallstensen, and E. Snekkenes, "A framework for estimating information security risk assessment method completeness," *Int. J. Inf. Secur.*, vol. 17, no. 6, pp. 681–699, Jun. 2017.
[16] *Information Technology—Security Techniques—Information Security Risk Management*, Standard ISO 27005–2011, 2011.
[17] J. Jones, "Factor analysis of information risk," U.S. Patent 10 912 863, Mar. 24, 2005,
[18] R. Caralli, J. F. Stevens, L. R. Young, and W. R. Wilson, "Introducing OCTAVE Allegro: Improving the information security risk assessment process," Softw. Eng. Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU/SEI-2007-TR-012, May 2007, p. 154. [Online]. Available: https://resources.sei.cmu.edu/asset_files/TechnicalReport/2007_005_001_14885.pdf
[19] Joint Task Force Transformation Initiative, "Guide for conducting risk assessments," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. NIST 800-30, 2012.
[20] ISACA. (2009). *The Risk IT Framework*. [Online]. Available: https://books.google.no/books?id=tG7VMihmwtsC
[21] F. Den Braber, I. Hogganvik, M. S. Lund, K. Stølen, and F. Vraalsen, "Model-based security analysis in seven steps—A guided tour to the CORAS method," *BT Technol. J.*, vol. 25, no. 1, pp. 101–117, 2007.
[22] L. Rajbhandari and E. Snekkenes, "Using the conflicting incentives risk analysis method," in *Security and Privacy Protection in Information Processing Systems* (IFIP Advances in Information and Communication Technology), vol. 405, L. J. Janczewski, H. B. Wolfe, and S. Shenoi, Eds. Berlin, Germany: Springer, 2013.
[23] S. H. Schwartz, "Basic human values: Theory, measurement and applications," *Revue française de Sociologie*, vol. 47, no. 4, p. 929, 2007.
[24] S. H. Schwartz, "An overview of the Schwartz theory of basic values," *Online Readings Psychol. Culture*, vol. 2, no. 1, p. 11, Dec. 2012.
[25] A. Szekeres, P. Wasnik, and E. Snekkenes, "Using demographic features for the prediction of basic human values underlying stakeholder motivation," in *Proc. 21st Int. Conf. Enterprise Inf. Syst.*, 2019, pp. 377–389.
[26] D. Difallah, E. Filatova, and P. Ipeirotis, "Demographics and dynamics of mechanical turk workers," in *Proc. 11th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2018, pp. 135–143.
[27] (2019), *Statistisk Sentralbyrå—Statistics Norway*. Accessed: Nov. 18, 2019. [Online]. Available: https://www.ssb.no/en
[28] S. H. Schwartz, "A proposal for measuring value orientations across nations," in *Proc. Questionnaire Package Eur. Social Surv.*, pp. 259–290, 2003.
[29] S. Schwartz. (2016). *Computing Scores for the 10 Human Values*. Accessed: Nov. 12, 2019. [Online]. Available: https://www.europeansocialsurvey.org/docs/methodology/ESS1_human_values_scale.pdf
[30] (2016). *Norwegian Version of PVQ-21*. [Online]. Available: https://www.europeansocialsurvey.org/docs/round8/fieldwork/norway/ESS8_questionnaires_NO.pdf/
[31] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *Int. J. Med. Edu.*, vol. 2, p. 53, 2011.
[32] N. Schmitt, "Uses and abuses of coefficient alpha," *Psychol. Assessment*, vol. 8, no. 4, pp. 350–353, 1996.
[33] R. Nisbet, J. Elder, and G. Miner, *Handbook of Statistical Analysis and Data Mining Applications*. New York, NY, USA: Academic, 2009.
[34] N. J. D. Nagelkerke, "A note on a general definition of the coefficient of determination," *Biometrika*, vol. 78, no. 3, pp. 691–692, 1991.
[35] T. Yarkoni and J. Westfall, "Choosing prediction over explanation in psychology: Lessons from machine learning," *Perspect. Psychol. Sci.*, vol. 12, no. 6, pp. 1100–1122, Aug. 2017.
[36] P. Bruce and A. Bruce, *Practical Statistics for Data Scientists: 50 Essential Concepts*. Newton, MA, USA: O'Reilly Media, 2017.
[37] S. H. Schwartz and A. Bardi, "Value hierarchies across cultures: Taking a similarities perspective," *J. Cross-Cultural Psychol.*, vol. 32, no. 3, pp. 268–290, Jul. 2016.
[38] O. Schwartz. (2019). *Untold History of Ai*. Accessed: Nov. 12, 2019. [Online]. Available: https://spectrum.ieee.org/tag/AI+history
[39] D. S. Char, N. H. Shah, and D. Magnus, "Implementing machine learning in health care—Addressing ethical challenges," *New England J. Med.*, vol. 378, no. 11, p. 981, 2018.
[40] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," 2017, *arXiv:1702.08608*. [Online]. Available: http://arxiv.org/abs/1702.08608
[41] C. Molnar. (2019). *Interpretable Machine Learning*. Accessed: Nov. 12, 2019. [Online]. Available: https://christophm.github.io/interpretable-ml-book/interpretability-imp%ortance.html

ADAM SZEKERES received the M.A. degree in cognitive psychology from the University of Szeged, Hungary. He is currently pursuing the Ph.D. degree in information security with NTNU Gjøvik, focusing on human motivation and the predictability of stakeholder behavior.

EINAR ARTHUR SNEKKENES received the Dr.Philos. degree in informatics from Oslo University, in 1995. He is currently a Full Professor with the Department of Information Security and Communication Technology, Norwegian University of Science and Technology (NTNU). His main research area is information security risk management.