Human Behavior Prediction for Risk Analysis

Adam Szekeres(✉) and Einar Snekkenes

Department of Information Security and Communication Technology,
Norwegian University of Science and Technology - NTNU,
Gjøvik, Norway
{adam.szekeres,einar.snekkenes}@ntnu.no

Abstract. The Conflicting Incentives Risk Analysis (CIRA) method makes predictions about human decisions to characterize risks within the domain of information security. Since traditional behavior prediction approaches utilizing personal features achieve low prediction accuracies in general, there is a need for improving predictive capabilities. Therefore, the primary objective of this study is to propose and test a psychological approach for behavior prediction, which utilizes features of situations to achieve improved predictive accuracy. An online questionnaire was used for collecting behavioral and trait data to enable a comparison of approaches. Results show that the proposed behavior prediction approach outperforms the traditional approach across a range of decisions. Additionally, interrater reliabilities are analyzed to estimate the extent of objectivity in situation evaluations, providing an indication about the potential performance of the approach when a risk analyst needs to rely on unobtrusive assessment of action-desirability.

Keywords: Risk analysis · Information security · Human motivation · Behavior prediction · Perception of situations

1 Introduction

Risk management aims to make predictions about potential future events, assess their consequences and mitigate undesirable outcomes by implementing appropriate controls. Within the Conflicting Incentives Risk Analysis (CIRA) method [25] risks result from conscious human decisions made by individuals. A decision-maker needs to consider how costs and benefits are allocated among the parties affected by the transaction: the decision-maker him/herself, other stakeholders, various groups and organizations. In the simplest case a decision has consequences only for the decision-maker. More often a personal choice impacts other stakeholders such as other individuals or an organization. In certain situations known as **threat risks** within CIRA personal benefits result in losses for other parties: Smart electricity meters have been reprogrammed for a fee by the employees of utility companies generating great financial losses for the organization [15]. Distribution system operators in the energy sector have to develop policies which balance the social costs of supply interruptions when creating policies for dealing with supply-demand imbalances [3]. Sexual harassment at workplaces causes trauma for the victims along with direct organizational costs (e.g. turnover and recruitment, investigating the complaint and legal penalties, damage to reputation paid by

The final authenticated version is available online at https://doi.org/10.1007/978-3-030-86970-0_21

the organization) [20]. In another class of situations, a personal cost (i.e. time, money, resources, freedom, etc.) has to be borne by an individual to provide a benefit for others. The lack of perceived incentives to act in a desirable manner gives rise to **opportunity risks** within CIRA: whistleblowers motivated by moral concerns take a significant personal risk for the benefit of society to uncover questionable practices within their organization [4]. An individual developing a web-service to provide free access to paywalled materials for everyone, risks personal freedom while publishers are faced with lost revenue [9]. Healthcare professionals have to deal with several costs (e.g. compliance with local and national regulations, delays in approval, complicated trial processes) to include patients in clinical trials for their benefit [13]. Spreading of malware can be prevented by a single individual sacrificing some resources (i.e. CPU, memory, storage) for the benefit of others. Risk management is becoming increasingly important as more and more domains of life become digitized and the number of people in interdependent relationships increases. Since organizational practices and legal systems identify individuals as responsible, accountable and punishable subjects for their actions [23,30], it is important to investigate how risks resulting from human decision-makers can be predicted at the individual level for the purpose of improving the CIRA method.

1.1 Problem Statement and Research Questions

To date, there is a lack of empirical tests investigating how a novel behavior prediction method performs within the CIRA method which takes into account differences in perception of situations among decision-makers. Thus, this study aims to empirically investigate the utility of a behavior prediction approach for improving the CIRA method, based on the following research questions: **RQ 1:** To what extent can predictive accuracy increase using an approach which takes into account differences in the perception of situational influences among decision-makers compared to a traditional approach utilizing personal features only? **RQ 2:** To what extent is the proposed method feasible when the situation perceptions need to be estimated unobtrusively by a risk analyst?

The paper is organized as follows: Sect. 2 presents existing results relevant to the paper's topics. Section 3 presents the characteristics of the sample, the instruments and procedures used for data collection as well as data processing steps. Results are presented in Sect. 4. A discussion of the results, their implications and the limitations of the study are discussed in Sect. 5. The key findings are summarized in Sect. 6 along with directions for further work.

2 Related Work

This section presents the risk analysis method under development; several types of risks in information security arising from conscious human behavior; and research results on the perception of situations providing the foundation for the proposed prediction approach.

2.1 Human Motivation at the Centre of Risk Analysis

The Conflicting Incentives Risk Analysis (CIRA) method developed within the domain of information security and privacy focuses on stakeholder motivation to characterize risks. The game-theoretic framework identifies two classes of stakeholders in an interdependent relationship [25]: **Strategy owners** are in position to select a certain strategy to implement an action based on the desirability of the available options, while **risk owners** are exposed to the actions or inactions of the strategy owners in a one shot game setting. Misaligned incentives can give rise to two types of risks: *threat risk* refers to undesirable outcomes for the risk owner, when the strategy owner selects a strategy based on its perceived desirability; *opportunity risk* refers to desirable outcomes for the risk owner, which fail to get realized, since the strategy owner has no incentives to act or would have to bear losses in order to provide a benefit for the risk owner. Motivation is represented by the expected change in overall utility after the execution of a strategy. The combination of several utility factors (i.e. personal attributes) contribute to each stakeholders' overall utility, and stakeholders are assumed to be utility maximisers. The method relies on the unobtrusive assessment of the strategy owner's relevant utility factors by a risk analyst, which is necessary since strategy owners are assumed to be adversarial (reluctant to interact with an analyst) and may be tempted to subvert the analysis by reporting false information. Previous work used the theory of basic human values (BHV) to operationalize a set of psychologically important utility factors [35], and investigated the extent to which observable features can be used for constructing the motivational profiles of strategy owners [37]. The theory of BHVs [29] distinguishes between 10 motivationally distinct desirable end-goals which form four higher dimensions as shown in Fig. 1. The trade-off among opposing values drives decision-making. Using a subject's profile information, the analyst makes a subjective assessment about action-desirability from the perspective of the strategy owner to predict future behavior [33].

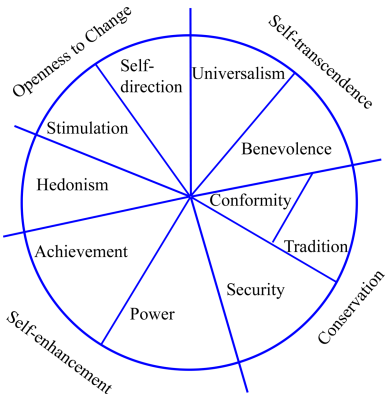

Fig. 1. Basic human value structure based on [29].

2.2 Risks Arising from Human Behavior Within Information Security

Insiders are trusted individuals who break information security policies deliberately. The purpose of a policy is to prescribe expected behaviors and to specify the consequences of undesirable behavior. Rule-breaking can be motivated by a variety of reasons (e.g. financial gain, curiosity, ideology, political, revenge) [19]. A recent systematic literature review revealed that most of the insider threat prediction applications focus on patterns of online activity as key features, whereas psychological approaches are less prevalent [6]. Some empirical work focuses on the psychological attributes of insiders, identifying and assessing personal or behavioral characteristics which are valid and reliable indicators of potential incidents. The insider threat prediction model prepared for the U.S. Department of Energy identifies 12 unique psychosocial behavioral indicators (e.g. disgruntlement, stress, absenteeism, etc.) that may be indicative of an insider threat [8]. The U.S. Department of Homeland Security [22] advises organizations to focus on a combination of personal attributes (e.g. introversion, greed, lack of empathy, narcissism, ethical flexibility) and behavioral indicators (e.g. using remote access, interest in matters outside of scope of duties, risk-taking behavior, etc.) to detect insider threats. Some of the most useful personal features for predicting insider threats are personality traits (e.g. Big Five, Dark Triad traits, sensation-seeking, etc.), emotions (e.g. hostility, anger) and mental disorders (e.g. paranoia, depression, etc.) [6].

Organizations are exposed to risks arising from negligent security practices by employees. Several empirical studies investigate conscious security-related behaviors using the Theory of Planned Behavior (TPB), which is one of the most widely used behavior prediction framework in psychology [1]. Using the theory (non-) compliant behavior can be predicted from the combination of the factors, and interventions can be targeted at specific factors to motivate desirable behavior. Most of the results included in [17] and [2] showed that all main constructs of TPB have significant associations with behavioral intentions, providing support for the model's utility. However, it is important to note that most studies measure behavioral intentions, as opposed to actual behavior. Prediction accuracy is generally measured by the R^2 metric averaging around $R^2 = 0.42$ across studies for intentions; but when actual behavior is measured accuracy may decrease to as low as $R^2 = 0.1$ [31], indicating that new models are needed to improve predictive capabilities.

A large number of risks are attributed to malicious external stakeholders. A taxonomy which categorizes hackers according to their properties (motivations, capabilities, triggers, methods) can be used for defense planning and forensic investigations. Key psychologically relevant motivations included in the taxonomy are: revenge, curiosity, financial gain, and fame [10].

2.3 Perception of Situations

The immediate situation in which decision-making takes place has important influence on behavior. It has been suggested, that behavior is a function of personal and situational attributes; known as the person-situation interactionist approach [18] and that the subject's perception of the situation determines their actions [14]. A simple process

model developed in [26] explains how subjects perceive and process situational information and generate different behavioral outcomes. Cues of the situation (e.g. persons, objects) are perceived by the subjects and processed according to their specific stable and variable personal features (e.g. traits, roles, mental states, etc.), giving rise to the unique experience of the situation. The actions selected are dependent on the experience, and two persons' action will match to the extent that they attend to the same cues of the environment, and process the cues similarly due to their similarities in terms of personal features. The literature of situations still lacks consensus on how to conceptualize, define and measure situations. This is mainly attributed to the complex and multifaceted nature of situations [21]. Attempts have been made to address this gap by developing taxonomies of situational features as perceived by the individual in the situation, assuming that behavioral incentives are subjective rather than objective [38]. Since the Dark Triad traits (narcissism, Machiavellianism, psychopathy) are frequently associated with harmful workplace behavior [34], a taxonomy was developed which identifies situational triggers contributing to the behavioral tendencies associated with these traits [24], which can be used for the development of situational interventions to mitigate risks. While descriptive situation taxonomies exist, applications for the purpose of behavior prediction are lacking.

3 Materials and Methods

This study used an online questionnaire for collecting data from subjects. Two types of behavioral data was collected using dilemmas which operationalized threat and opportunity risks: **subjective ratings** capturing perception of situations; and **explicit choices** among the two options of each dilemma. Additionally, the following personal features were collected: basic demographic information and motivational profiles operationalized as BHVs. Sections were presented in the following order to maximize the number of questions between the two behavioral tasks:

1. Perception of situations (i.e. subjective ratings of dilemma-options).
2. Basic demographic information and motivational profiles (i.e. BHVs).
3. Choice between two options of each dilemma.

The questionnaire was completely anonymous, no personally identifiable information was collected, participants were required to express consent before starting the questionnaire. The questionnaire was implemented in Limesurvey and was hosted on servers provided by the Norwegian University of Science and Technology (NTNU).

3.1 Sample

Relying on the sample size recommendations for logistic regression analyses, data collection aimed at a minimum of 50 fully completed questionnaires [39]. First, a randomly selected sample of university students received an e-mail invitation to fill the online questionnaire, generating 22 fully completed surveys. Next, 40 additional respondents were recruited through the Amazon Mechanical Turk (MTurk) online workplace, where

subjects receive payment for completing various human intelligence tasks (HITs). Each respondent who completed the questionnaire received $4 net compensation through the MTurk system, which equals to an hourly rate of $12–16. In addition to the higher-than average payment [12], additional options were selected to ensure data quality: the survey was available only for MTurk workers with a HIT Approval Rate greater than 90%, and only to Masters (using MTurk's quality assurance mechanism). Questionnaires below 9 min of completion time were entirely removed to increase the quality of the dataset. Thus, the final convenience sample contains data from 59 respondents (27 females and 32 males) with a mean age of 34 years (S.D. = 10.44). Citizenship of the respondents is as follows: 53% U.S., 25% Norway, 14% India, 8 % other. Most respondents have bachelor's degree (46%), followed by a completed upper secondary education (36%), master's degree (17%) and lower secondary education (2%).

3.2 Behavioral Data

Dilemmas. The stimuli for collecting two types of behavioral data (i.e. perception of situations and explicit choices) were dilemmas which operationalized threat risks as moral dilemmas and opportunity risks as altruistic dilemmas, based on a taxonomy of situations [36]. The dilemmas are based on real events and cover a broad range of behaviors resulting in threat or opportunity risks to the specified risk owners to increase the ecological validity of the stimuli. Dilemmas were presented as riskless choices (i.e. consequences are specified with certainty). Table 1 provides a summary about the theme of the nine dilemmas used in the questionnaire and the first dilemma is presented in full detail in the Appendix due to space limitations. Each dilemma comprised of a story (setting the context as shown in Item 1), and two mutually exclusive options as presented in Item 2 and Item 3 in the Appendix. Respondents answered the questions as strategy

Table 1. Short description of the main theme of the dilemmas included in the survey.

No	Theme of dilemmas
1	Kill an injured person to save rest of crew?*
2	Approach employees with a sexual offer looking for promotion?
3	Distribute electricity to residents instead of hospital during electricity crisis?
4	Reprogram customer's Smart Meters for a personal gain against the rules?
5	Inform customers of your employer about security issues identified in your products despite sure prosecution?
6	Include a patient in clinical trial despite significant personal costs?
7	Create paywall bypassing website to make research results freely available at the expense of personal freedom?
8	Sacrifice personal resources to run a virus to protect colleagues?
9	Accept Firm B's offer? Firm A: $100.000 salary + 14 days holiday vs. Firm B: $50.000 salary + 16 days holiday.*

Note. Dilemmas marked with * were taken from [7]

owners. The dilemmas were used at the beginning of the questionnaire to collect subjective ratings about the perception of situations, and at the end of the questionnaire to collect explicit choices from participants (explained below).

Perception of Situations. To capture the perceived losses/benefits from a particular choice, subjects were required to rate separately both options of all dilemmas (see Item 2 and Item 3) along five dimensions as shown in Fig. 4 in the Appendix. Ratings were collected by continuous sliding scales ranging from negative 100 through 0 to positive 100, with textual anchor labels at the two endpoints and at the mid-point of the scale (i.e. −100: Maximum possible decrease; 0: No impact; 100: Maximum possible increase). For each option an overall utility score was calculated as follows: $U_{\text{total}} = U_{\text{socialstatus}} + U_{\text{careforothers}} + U_{\text{excitement}} + U_{\text{stability}} + U_{\text{pleasure}}$, using an unweighted version of the multi-attribute utility theory (MAUT) [5] implemented in CIRA [25]. This step enables the comparison of the desirability of both options and provides a basis for checking the internal consistency of choices. A choice was considered internally consistent when the explicitly chosen option (in Section 3 of questionnaire) received a higher utility score than the unchosen option of the same dilemma based on ratings (Section 1 of the questionnaire). This metric gives an indication of data validity (i.e. whether respondents were following instructions) and subject rationality (i.e. making choices according to stated preferences).

Choice Between Dilemma-Options. Figure 5 in the Appendix shows an example of the final section of the questionnaire, requiring participants to make an explicit choice between two dilemma options as strategy owners. The proposed prediction method uses the subjective ratings (Section 1 of questionnaire) corresponding to the chosen option (Section 3 of questionnaire) for each subject for all dilemmas. For example, if a subject selected *option NO* in the choice task on *dilemma_1*, the ratings for *option NO* of *dilemma_1* were used as predictors, even if the total utility was higher for *option YES* of *dilemma_1*.

3.3 Personal Features: Demographics and Motivational Profile

Basic demographic data included age, sex, nationality, level of education. Individual motivational profiles were created using the Portray Value Questionnaire (PVQ-21), which is a 21-item questionnaire designed for self-assessment [28]. The instrument captures ten BHVs, which were computed according to the instructions in [28]. Five higher dimensions were created by computing the mean of the corresponding values as follows: *Self-enhancement:* power and achievement, *Self-transcendence:* universalism and benevolence, *Openness to change:* self-direction and stimulation, *Conservation:* security, tradition, conformity, while *Hedonism* was treated as a separate dimension, since "hedonism shares elements of both Openness and Self-Enhancement" [27].

4 Results

All analyses were conducted in IBM SPSS 25. Figure 2 presents descriptive statistics for all dilemmas. Red bars indicate the percentage of affirmative choices provided by

subjects in the third section of the questionnaire. Blue bars represent the percentage of internally consistent choices. Dilemma 9 was included as a control question. It received a high number of internally consistent (signifying that evaluations provided by subjects were valid), correct responses (selecting the option with higher utility), indicating that most of the respondents were following instructions properly, thus results can be considered valid.

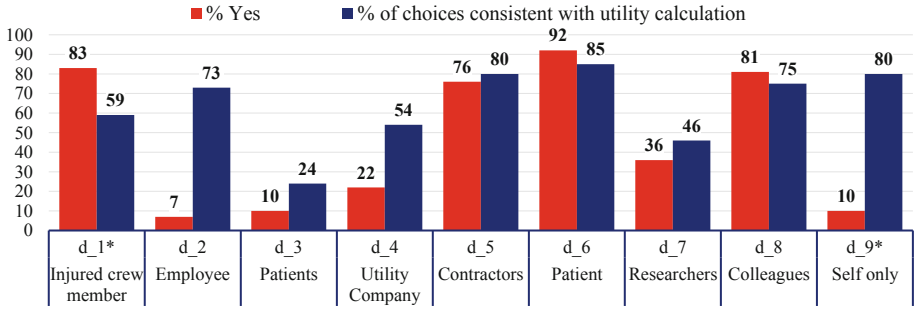

Fig. 2. Descriptive statistics for nine dilemmas. Percentage of affirmative choices and percentage of choices consistent with utility calculations derived from subjective situation ratings across dilemmas. Risk owners are identified under the dilemma numbers. The YES options refer to threat risks realized and opportunity risks avoided for the risk owners. Threat risks: d_1–d_4; Opportunity risks: d_5–d_8. A rational dilemma (d_9) was included as a control question. Dilemmas marked with * were taken from [7].

4.1 RQ 1: Comparison of Prediction Approaches

In order to assess the differences between the traditional (i.e. utilizing personal features only) and the proposed (i.e. utilizing perception of situations) approaches two separate sets of analyses were conducted. First, nine binary logistic regression models were built (one for each dilemma) using the motivational profiles as independent variables and explicit choices as dependent variables. The second set of analyses aimed at exploring the extent of potential improvements that can be expected when subjective perceptions of situations are used to predict the same choices. In the first case five of the nine predictive models were significantly better than the intercept-only models. In case of the proposed approach seven predictive models were significantly better than the intercept-only models. Table 2 presents a summary of all the logistic regression models' predictive performance. The proposed approach ("Subjective ratings of situation" columns) outperformed the traditional approach ("Personal features only" columns) in seven cases out of nine total cases, according to all performance metrics (percentage of correct classifications and the Nagelkerke R^2 - total variance explained).

Table 3 and Table 4 in Appendix presents the details for each model with the regression coefficients and corresponding tests of significance for each predictor. Predictive performance for each model is assessed by two variants of the R^2 metric: Cox & Snell

Table 2. Comparison of the two approaches for predicting identical outcomes.

	% of overall correct classification		% of variance explained (Nagelkerke's R^2)	
	Personal features only	Subjective ratings of situation	Personal features only	Subjective ratings of situation
Dilemma_1	86.4	**89.8***	30	**50***
Dilemma_2	93.2	91.5	44	41
Dilemma_3	86.4	**94.9***	36	**58***
Dilemma_4	74.6	**81.4***	24	**50***
Dilemma_5	76.3	**91.5***	11	**70***
Dilemma_6	89.8	93.2	28	31
Dilemma_7	71.2	**91.5***	29	**76***
Dilemma_8	81.4	**86.4***	30	**45***
Dilemma_9	91.5	**93.2***	35	**40***

Note. * improvement in predictive accuracy from traditional approach for models which were significantly better than the intercept-only model

R^2 and Nagelkerke R^2, measuring the total variances explained by the models. The "Overall model evaluation" row of the tables show which models performed better than the intercept-only models.

4.2 RQ 2: Practical Feasibility of the Proposed Predictive Approach

The second research question is concerned with exploring the potential accuracy which can be expected when ratings must be assessed unobtrusively by a risk analyst (i.e. subjective evaluations are not available). The level of agreement about the ratings between subjects indicates the level of objectivity in situation perceptions. In order to analyze the extent of objectivity, data was prepared as follows: as dilemma-options represent independent objects which were rated by subjects, for each dilemma-option (18 in total) a separate dataset was created, thus dilemma-options represent the unit of analysis. Each respondent was entered in the columns and situation ratings were entered as rows in each dataset following the guidelines in [32]. Intraclass correlations (ICC) are used as estimates of inter-rater reliability, a technique which is useful for understanding the proportion of reliable ("real") estimates provided by independent raters about a construct or a combination of constructs [16]. As respondents represent a sample from the population of potential respondents and all dilemma options were evaluated by all raters a two-way random analysis was selected ICC(2), which assumes random effects for raters as well as for the constructs being rated. Two types of reliability scores (consistency and absolute agreement) were computed to assess the accuracies of the ratings using ICC(2, 1). The difference between the consistency and absolute agreement measures is that "if two variables are perfectly consistent, they don't necessarily agree" [16]. Absolute agreement represents a more restrictive measure of inter-rater reliability. Figure 3 presents the intraclass correlation scores for all dilemmas. Consistency (red bars) refers to the extent of agreement about the directions of the ratings using a randomly selected analyst. Absolute agreement (blue bars) represents the expected accuracy when a single analyst estimates the exact scores. "An interrater reliability estimate of 0.80 would indicate that 80% of the observed variance is due to true score variance or similarity in ratings between coders, and 20% is due to error variance or differences in ratings between coders" [11].

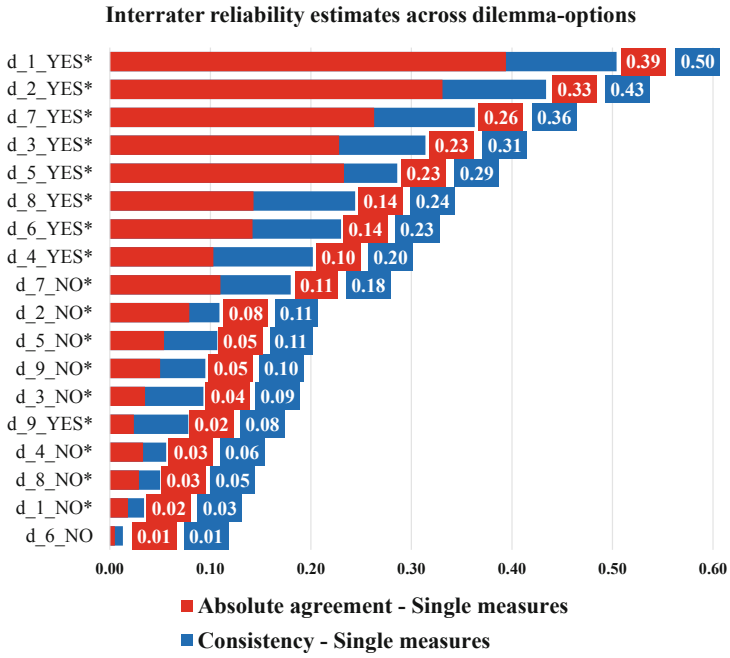

Fig. 3. Interrater reliability estimates across all dilemma-options. Red bars indicate the consistency of ratings, blue bars represent absolute agreement. Dilemma-options marked with * are statistically significant at $p \leq 0.05$. (Color figure online)

5 Discussion

Predictability is the essence of security. The primary purpose of the study was to test a proof of concept related to human behavior prediction using a novel approach. The first research question focused on assessing the extent of predictive improvement when subjective situation perceptions are used as predictors compared to a traditional prediction approach (personal features only). Two sets of analyses were conducted to enable an unbiased comparison between the two approaches using the same number of predictors. The overall percentage of correct classifications ranges between 71.2%–93.2% when only personal features are utilized, and between 81.4%–94.9% using the proposed approach. Predictive performance in terms of Nagelkerke's R^2 performance metric consolidates the findings for the proposed method's superiority. Nagelkerke's R^2 scores for the traditional models range between: 11%–44%, while for the proposed approach, predictive performance ranges between: 31%–76%. In sum, the proposed approach is able to predict more choices, as well as the models based on the proposed approach explain more variance in the outcomes.

The second research question aimed at exploring the extent of objectivity inherent in situation perceptions. The question is relevant when a risk analyst has to unobtrusively estimate the perceptions, due to unavailability of subjects. Intraclass correlations were used as estimates of interrater reliability to explore the extent of agreement between subjects about situation perceptions. Absolute agreement ranges between 2–39% across dilemmas, while consistency ranges between 3–50% showing a significant extent of subjectivity in perceptions, necessitating further work to reduce uncertainty.

Limitations can be identified in terms of the composition of the sample and the ecological validity of the study. Due to difficulties with subject recruitment, a convenience sample was used, which limits the generalizability of the results. Threat risks and opportunity risks are complex, emergent properties resulting from the interaction between the stakeholders. The dilemmas developed for this study aimed at capturing these risk types and were presented as hypothetical stories to respondents. Even though self-report questionnaires are the most-widely used methods for data collection, they may be prone to several problems (e.g. socially desirable responding, lack of experimental control, etc.) hampering the validity of the results. On the other hand, the anonymous nature of the online questionnaire can facilitate the expression of socially undesirable intentions, increasing overall validity.

6 Conclusions and Further Work

Risks attributed to human decisions are increasingly prevalent in all domains of life, necessitating better models and predictive capabilities for risk analysis. Therefore, this study aimed at contributing to the field by enhancing the predictive capabilities of the CIRA method, which focuses on decision-makers' motivation when characterizing risks. A behavior prediction approach was proposed and evaluated, which relies on how situations are perceived by decision-makers. The proposed approach consistently outperforms the traditional approach, but further work is needed to improve its utility in settings where unobtrusive assessment is the only option for a risk analyst. Future studies need to explore whether it is possible to increase the accuracy of situation perceptions by observers. This could be achieved by training analysts in situation-assessment. Furthermore, the development of automated situation-assessments would be necessary to increase reliability. The present study used dilemmas in which "the decision's impact on other people" can be considered the most salient feature of the environment, thus future studies could explore how this salience impacts decision-makers in real-world organizations.

Appendix

1. **Story of dilemma_1:** You are the captain of a military submarine travelling underneath a large iceberg. An onboard explosion has caused you to lose most of your oxygen supply and has injured one of your crew who is quickly losing blood. The injured crew member is going to die from his wounds no matter what happens. The remaining oxygen is not sufficient for the entire crew to make it to the surface. The only way to save the other crew members is to shoot dead the injured crew member so that there will be just enough oxygen for the rest of the crew to survive.
2. **NO option of dilemma_1:** By considering the consequences please rate how **NOT shooting the injured crew member** would influence each of the following factors from your perspective compared to your state before the decision!
3. **YES option of dilemma_1:** By considering the consequences please rate how **shooting the injured crew member** would influence each of the following factors from your perspective compared to your state before the decision!

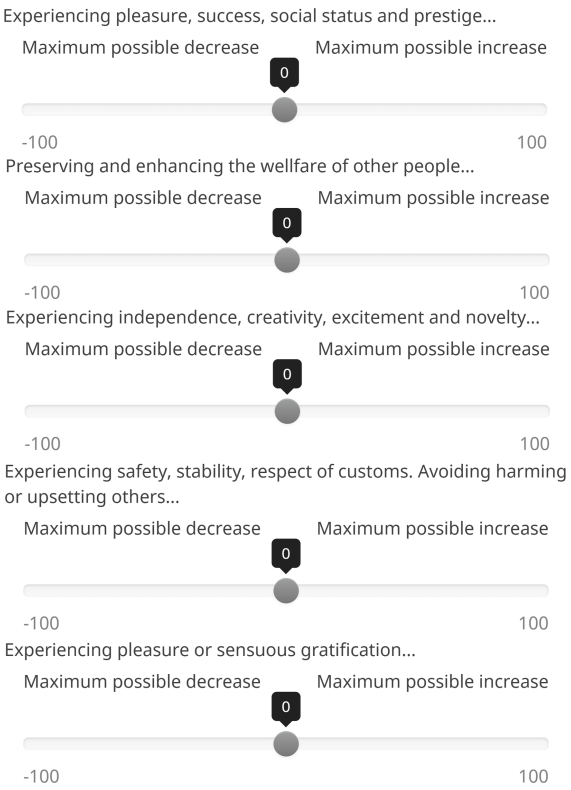

Fig. 4. The rating scales used to collect the subjective perceptions about each situation (i.e. dilemma-option).

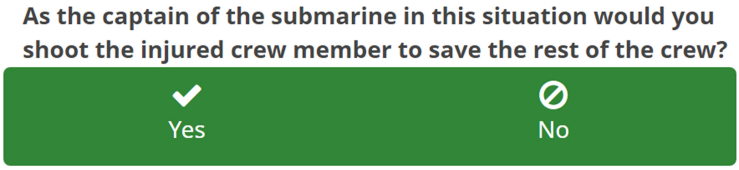

Fig. 5. Task requiring an explicit choice between the two options of dilemma_1.

Table 3. Summary of nine binary logistic regression models for each dilemma. Each model uses personal features (BHVs) of subjects as independent variables to predict choices.

Predictor		d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
Constant	β	9.04	−12.83	−5.53	−4.39	−1.67	7.74	−7.94	5.58	−10.72
	SE β	4.01	7.42	3.34	2.54	2.14	4.64	2.86	3.31	5.30
	p	**0.02***	0.08	0.10	0.09	0.44	0.10	**0.01***	0.09	**0.04***
Self-enhancement	β	−0.36	3.12	1.31	0.44	−0.02	−0.61	0.45	−1.48	1.01
	SE β	0.65	1.90	0.92	0.52	0.45	0.91	0.44	0.64	0.87
	p	0.58	0.10	0.15	0.40	0.97	0.50	0.30	**0.02***	0.24
Self-transcendence	β	−0.62	−0.47	−1.13	−0.75	0.67	1.18	0.17	0.11	−0.51
	SE β	0.74	1.34	0.79	0.60	0.52	1.11	0.58	0.79	1.11
	p	0.40	0.73	0.15	0.21	0.20	0.29	0.76	0.89	0.65
Openness to change	β	0.53	−2.18	−0.63	−0.01	0.39	−1.03	0.33	−0.26	0.83
	SE β	0.71	1.80	0.93	0.62	0.49	1.25	0.53	0.76	1.14
	p	0.46	0.23	0.50	0.99	0.43	0.41	0.54	0.73	0.47
Conversation	β	−1.49	1.43	0.30	1.00	−0.06	−1.06	0.71	−0.19	0.49
	SE β	0.67	1.45	0.71	0.52	0.40	0.82	0.42	0.47	0.66
	p	**0.03***	0.32	0.67	**0.05***	0.87	0.19	0.09	0.69	0.46
Hedonism	β	0.27	1.11	1.38	0.32	−0.39	0.01	0.25	0.43	0.46
	SE β	0.39	0.98	0.74	0.39	0.36	0.66	0.34	0.48	0.67
	p	0.49	0.26	0.06	0.41	0.27	0.99	0.47	0.37	0.49
Overall model evaluation	χ^2	11.80	11.14	11.15	10.08	4.51	7.86	14.17	11.89	10.90
	df	5								
	p	**0.04***	**0.05***	**0.05***	0.07	0.48	0.16	**0.02***	**0.04***	0.06
Goodness-of-fit-tests:										
Cox and Snell R^2		0.18	0.17	0.17	0.16	0.07	0.13	0.21	0.18	0.17
Nagelkerke R^2		0.30	0.44	0.36	0.24	0.11	0.28	0.29	0.30	0.35

Note. ***p** $\leq$ **0.05.**

Table 4. Summary of nine binary logistic regression models for each dilemma. Each model uses subjective ratings of situations as independent variables to predict choices.

Predictor		d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
Constant	β	0.90	−4.66	−2.87	−2.41	−0.63	0.14	−1.46	1.04	−3.31
	SE β	0.64	1.72	0.90	0.66	0.72	0.96	0.60	0.48	1.11
	p	0.16	**0.01***	**0.00***	**0.00***	0.38	0.89	**0.02***	**0.03***	**0.00***
S_1 Social status	β	−0.04	0.02	−0.01	0.01	−0.02	−0.02	0.02	−0.02	−0.06
	SE β	0.01	0.02	0.02	0.02	0.02	0.03	0.03	0.02	0.03
	p	**0.01***	0.34	0.45	0.75	0.40	0.55	0.48	0.23	**0.03***
S_2 Care for others	β	0.02	0.00	0.04	0.03	0.06	0.05	−0.03	0.02	0.03
	SE β	0.01	0.02	0.02	0.01	0.02	0.03	0.03	0.01	0.02
	p	0.08	0.99	**0.02***	**0.02***	**0.00***	0.10	0.26	0.12	0.23
S_3 Excitement	β	−0.02	0.06	0.00	−0.01	0.00	0.01	0.06	0.00	0.08
	SE β	0.01	0.03	0.02	0.02	0.03	0.02	0.03	0.01	0.03
	p	0.32	**0.05***	0.97	0.43	0.99	0.81	**0.01***	0.91	**0.01***
S_4 Stability	β	0.02	−0.03	0.01	−0.03	0.00	0.01	−0.03	0.03	0.00
	SE β	0.01	0.02	0.02	0.01	0.02	0.02	0.02	0.01	0.02
	p	0.12	0.19	0.48	**0.03***	0.82	0.68	**0.03***	**0.03***	0.96
S_5 Pleasure	β	0.01	−0.02	−0.01	0.04	0.00	0.01	0.02	−0.02	−0.02
	SE β	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.02
	p	0.36	0.37	0.69	**0.04***	0.92	0.65	0.29	0.22	0.28
Overall model evaluation	χ^2	21.00	10.40	19.45	22.97	37.13	8.50	47.65	19.15	12.64
	df	5								
	p	**0.00***	0.07	**0.00***	**0.00***	**0.00***	0.13	**0.00***	**0.00***	**0.03***
Goodness-of-fit-tests:										
Cox and Snell R^2		0.30	0.16	0.28	0.32	0.47	0.13	0.55	0.28	0.19
Nagelkerke R^2		0.50	0.41	0.58	0.50	0.70	0.31	0.76	0.45	0.40

Note. ***p* ≤ *0.05.*** *S_1-5: subjective ratings of situational features collected in section* 1 *of the questionnaire.*

References

1. Ajzen, I.: The theory of planned behavior. Organ. Behav. Hum. Decis. Process. **50**(2), 179–211 (1991)
2. Cram, W.A., Proudfoot, J., D'Arcy, J.: Seeing the forest and the trees: a meta-analysis of information security policy compliance literature. In: Proceedings of the 50th Hawaii International Conference on System Sciences, pp. 4051–4060 (2017)
3. De Nooij, M., Lieshout, R., Koopmans, C.: Optimal blackouts: empirical results on reducing the social cost of electricity outages through efficient regional rationing. Energy Econ. **31**(3), 342–347 (2009)
4. Dungan, J.A., Young, L., Waytz, A.: The power of moral concerns in predicting whistleblowing decisions. J. Exp. Soc. Psychol. **85**, 103848 (2019)
5. Fischer, G.W.: Experimental applications of multi-attribute utility models. In: Wendt D., Vlek C. (eds.) Utility, Probability, and Human Decision Making. Theory and Decision Library. An International Series in the Philosophy and Methodology of the Social and Behavioral Sciences, vol. 11. Springer, Dordrecht (1975). https://doi.org/10.1007/978-94-010-1834-0_
6. Gheyas, I.A., Abdallah, A.E.: Detection and prediction of insider threats to cyber security: a systematic literature review and meta-analysis. Big Data Anal. **1**(1), 6 (2016)

7. Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen, J.D.: An fMRI investigation of emotional engagement in moral judgment. Science **293**(5537), 2105–2108 (2001)
8. Greitzer, F.L., Kangas, L.J., Noonan, C.F., Dalton, A.C.: Identifying at-risk employees: a behavioral model for predicting potential insider threats. Technical note PNNL-19665, Pacific Northwest National Lab (PNNL), Richland, WA (2010)
9. Greshake, B.: Looking into Pandora's Box: the content of sci-hub and its usage. F1000Research **6**, 541 (2017)
10. Hald, S.L., Pedersen, J.M.: An updated taxonomy for characterizing hackers according to their threat properties. In: 2012 14th International Conference on Advanced Communication Technology (ICACT), pp. 81–86. IEEE (2012)
11. Hallgren, K.A.: Computing inter-rater reliability for observational data: an overview and tutorial. Tut. Quant. Meth. Psychol. **8**(1), 23 (2012)
12. Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., Bigham, J.P.: A data-driven analysis of workers' earnings on amazon mechanical turk. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2018)
13. James, S., Rao, S.V., Granger, C.B.: Registry-based randomized clinical trials-a new clinical trial paradigm. Nat. Rev. Cardiol. **12**(5), 312–316 (2015)
14. Kihlstrom, J.F.: The person-situation interaction. In: The Oxford Handbook of Social Cognition, pp. 786–805. Oxford University Press (2013)
15. Krebs, B.: FBI: Smart Meter Hacks Likely to Spread—Krebs on Security (April 2012). https://krebsonsecurity.com/2012/04/fbi-smart-meter-hacks-likely-to-spread. Accessed 14 Feb 2020
16. Landers, R.N.: Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. Winnower **2**, e143518 (2015)
17. Lebek, B., Uffen, J., Neumann, M., Hohler, B., Breitner, M.H.: Information security awareness and behavior: a theory-based literature review. Manage. Res. Rev. **37**, 1049–1092 (2014)
18. Lewin, K.: Principles of Topological Psychology. McGraw-Hill (1936)
19. Maasberg, M., Warren, J., Beebe, N.L.: The dark side of the insider: detecting the insider threat through examination of dark triad personality traits. In: 2015 48th Hawaii International Conference on System Sciences, pp. 3518–3526. IEEE (2015)
20. McDonald, P.: Workplace sexual harassment 30 years on: a review of the literature. Int. J. Manag. Rev. **14**(1), 1–17 (2012)
21. Meyer, R.D.: Taxonomy of Situations and Their Measurement (2015)
22. N.A.: Combating the insider threat. Technical report, Department of Homeland Security (May 2014). https://www.us-cert.gov/sites/default/files/publications/Combating%20the%20Insider%20Threat_0.pdf
23. Naffine, N.: Who are law's persons? From Cheshire cats to responsible subjects. Mod. Law Rev. **66**(3), 346–367 (2003)
24. Nübold, A., et al.: Developing a taxonomy of dark triad triggers at work-a grounded theory study protocol. Front. Psychol. **8**, 293 (2017)
25. Rajbhandari, L., Snekkenes, E.: Using the conflicting incentives risk analysis method. In: Janczewski, L.J., Wolfe, H.B., Shenoi, S. (eds.) SEC 2013. IAICT, vol. 405, pp. 315–329. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39218-4_24
26. Rauthmann, J.F., Sherman, R.A., Funder, D.C.: Principles of situation research: towards a better understanding of psychological situations. Eur. J. Pers. **29**(3), 363–381 (2015)
27. Schwartz, S.: Value priorities and behavior: applying a theory of intergrated value systems. In: The Psychology of Values: The Ontario Symposium, vol. 8, pp. 119–144 (2013)
28. Schwartz, S.: Computing scores for the 10 human values (2016). https://www.europeansocialsurvey.org/docs/methodology/ESS_computing_human_values_scale.pdf. Accessed 02 Mar 2021

29. Schwartz, S.H.: An overview of the Schwartz theory of basic values. Online Read. Psychol. Cult. **2**(1), 1–20 (2012)
30. Selznick, P.: Foundations of the theory of organization. Am. Sociol. Rev. **13**(1), 25–35 (1948)
31. Shropshire, J., Warkentin, M., Sharma, S.: Personality, attitudes, and intentions: predicting initial adoption of information security behavior. Comput. Secur. **49**, 177–191 (2015)
32. Shrout, P.E., Fleiss, J.L.: Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. **86**(2), 420 (1979)
33. Snekkenes, E.: Position paper: privacy risk analysis is about understanding conflicting incentives. In: Fischer-Hübner, S., de Leeuw, E., Mitchell, C. (eds.) IDMAN 2013. IAICT, vol. 396, pp. 100–103. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37282-7_9
34. Spain, S.M., Harms, P., LeBreton, J.M.: The dark side of personality at work. J. Organ. Behav. **35**(S1), S41–S60 (2014)
35. Szekeres, A., Snekkenes, E.A.: Predicting CEO misbehavior from observables: comparative evaluation of two major personality models. In: Obaidat, M.S. (ed.) ICETE 2018. CCIS, vol. 1118, pp. 135–158. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-34866-3_7
36. Szekeres, A., Snekkenes, E.A.: A taxonomy of situations within the context of risk analysis. In: Proceedings of the 25th Conference of Open Innovations Association FRUCT, pp. 306–316. FRUCT Oy. Helsinki, Finland (2019)
37. Szekeres, A., Snekkenes, E.A.: Construction of human motivational profiles by observation for risk analysis. IEEE Access **8**, 45096–45107 (2020)
38. Ten Berge, M.A., De Raad, B.: Taxonomies of situations from a trait psychological perspective: a review. Eur. J. Pers. **13**(5), 337–360 (1999)
39. VanVoorhis, C.W., Morgan, B.L.: Understanding power and rules of thumb for determining sample sizes. Tut. Quant. Meth. Psychol. **3**(2), 43–50 (2007)