Using Internet Activity Profiling for Insider-threat Detection

Bushra A. Alahmadi[1], Philip A. Legg[2] and Jason R. C. Nurse[1]
[1]*Cyber Security Centre, Department of Computer Science, University of Oxford, Oxford, U.K.*
[2]*Department of Computer Science, University of the West of England, Bristol, U.K.*

Keywords: Insider-threat Detection, Behavioural Analysis, Internet Activity, Psychological Traits.

Abstract: The insider-threat problem continues to be a major risk to both public and private sectors, where those people who have privileged knowledge and access choose to abuse this in some way to cause harm towards their organisation. To combat against this, organisations are beginning to invest heavily in deterrence monitoring tools to observe employees' activity, such as computer access, Internet browsing, and email communications. Whilst such tools may provide some way towards detecting attacks afterwards, what may be more useful is preventative monitoring, where user characteristics and behaviours inform about the possibility of an attack before it happens. Psychological research advocates that the behaviour and preference of a person can be explained to a great extent by psychological constructs called personality traits, which could then possibly indicate the likelihood of an individual being a potential insider threat. By considering how browsing content relates to psychological constructs (such as OCEAN), and how an individual's browsing behaviour deviates over time, potential insider-threats could be uncovered before significant damage is caused. The main contribution in this paper is to explore how Internet browsing activity could be used to predict the individual's psychological characteristics in order to detect potential insider-threats. Our results demonstrate that predictive assessment can be made between the content available on a website, and the associated personality traits, which could greatly improve the prospects of preventing insider attacks.

1 INTRODUCTION

Insider threats continue to be a paramount cyber security challenge that threaten businesses, institutions, and governmental organisations, a risk highlighted prominently after Snowden's 2011 revelations. Malicious insiders are current or former employees or trusted partners within the organisation perimeter that abuse their authorised access to the organisation network or system (Hunker and Probst, 2011). Academic literature and industry surveys and reports provide unequivocal evidence that support the significant impact caused by insider threats (Cappelli et al., 2012).

In the aim of deterring such a threat, organisations have defined technological monitoring mechanisms that facilitate the tracking of employees' computer activities such as file access (Ehrman et al., 1997), email communications (Nord et al., 2006), Internet browsing patterns (Urbaczewski and Jessup, 2002) and host/network activity monitoring such as intrusion detection or intrusion prevention devices and honey-pots (Spitzner, 2003). However, some of these approaches require a reactive response whilst others are simply inadequate due to the fact that insiders have authorised access to organisation assets and data. Researchers argue that a more comprehensive analysis approach that incorporates a diverse set of data sources from technological monitoring profiling to behavioural and psychological observations is more effective in accurate recognition, detection and response to insider threats (Schultz, 2002)(Greitzer and Frincke, 2010)(Legg et al., 2013)(Nurse et al., 2014). Such a modelling approach would flag high risk and concerning activities on which organisation's cyber-security analysts can focus their attention, without having to individually assess the full complement of data gathered. For example, it could link downloading a huge file with an individual's behavioural profile such as work stress or financial problems, thus providing another dimension amongst which to assess a potential malicious act that might have been ignored or missed in traditional monitoring approaches.

Psychology research advocates that an individual's behaviour and preferences can be explained to a great extent by psychological constructs called personality traits (Allport, 1962). This implies that knowledge of an individual's personality enables some level of prediction of behaviour and preferences

Alahmadi B., Legg P. and Nurse J..
Using Internet Activity Profiling for Insider-threat Detection.
DOI: 10.5220/0005480407090720
In *Proceedings of the 17th International Conference on Enterprise Information Systems* (WOSIS-2015), pages 709-720
ISBN: 978-989-758-097-0

across different contexts and environments (Kosinski et al., 2014). Such personality characterisations include OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) (Wiggins, 1996) and the Dark Triad (Machiavellianism, Narcissism, and Psychopathy) (Paulhus and Williams, 2002). As personality traits influence an individual's actions, they are also a key factor in whether or not the individual is likely to carry out a malicious insider attack (Legg et al., 2013). Although personality traits and behaviour alone are not an indicative of an individual's insider threat potential, when combined with other observations this can yield significant confidence in the threat posed.

The high frequency of web browsing usage provide an opportunity for organisations to detect a potential insider threat through analysis of employees' browsing interests. Similar to an individual's preference in movies or literature, browsing patterns can allude to an individual's likes, dislikes, and personal interests (Shaban et al., 2010). In addition, previous research have succeeded in predicting individuals' personality cues by analysing their social media accounts or personal websites (Sumner et al., 2012) (Golbeck et al., 2011b) (Shen et al., 2013). However, to the extent of our knowledge, no research has been done so far that aims to predict the personality traits of an individual based on their browsing interests.

In this paper, we examine how personality is manifested in individuals' browsing interests. We hypothesise that given a large collection of a user's web browsing history, one could infer preferences and traits related to personality. As a result, deviations in such behaviour can signify a change in personality which can characterise insider threat. Such an approach could be utilised by organisations to monitor their employees to detect any sudden deviation that could potentially indicate an increased likelihood of insider attack. This information could be utilised for insider-threat detection in a comprehensive model (Schultz, 2002)(Greitzer and Frincke, 2010)(Legg et al., 2013)(Nurse et al., 2014). With respect to insider-threat detection, we specifically present an initial approach to:

- Map website keywords to OCEAN personality traits; and
- Using the keyword-personality mapping create a browsing profile that could be monitored to detect browsing deviations and consequently using these deviations to uncover changes in personality traits.

In particular, we provide the following research contributions:

- Explore the feature extraction techniques for web content analysis;
- Performance analysis of web content correlations in conjunction with psychological traits;
- Identify relationships between websites and psychological traits; and
- Explore how psychological traits and website correlations can aid in insider-threat detection.

The organisation of the rest of the paper is as follows. In Section 2, we discuss related work on personality traits, their prediction and their utility at assisting in the detection of insider threats. Section 3 defines our approach to correlate personality traits with browsing interests, and how these inferred characteristics may be indicative of an insider threat. In Section 4, we present our experimental setup and methods for analysing textual features of websites and their correlation with personality traits. We present the results of the correlations between each website choice and personality trait in Section 5. In Section 6, we discuss the experiment results and the implications that this work has for organisations that may utilise website browsing preferences to better predict potential insider-threat risks. In Section 7, we conclude and discuss future work.

2 RELATED WORK

Academic research and industry reports have emphasised the greater threat that insiders pose and proposed various deterrence and monitoring mechanisms (Cappelli et al., 2012)(PWC, 2014). Although traditional insider-threat monitoring tools serve their use, several papers have advocated the need for a comprehensive framework that combines technical monitoring mechanisms with behavioural observations and human resources information for more robust insider-threat detection (Schultz, 2002)(Greitzer and Frincke, 2010)(Legg et al., 2013)(Nurse et al., 2014). Such contributions include defining an insider-threat framework that considers various metrics such as psychological profiling and behaviour patterns. For example, Phyo and Furnell defined an approach for dealing with insider threats that result from IT system abuse by profiling user behaviour (Phyo and Furnell, 2004).

Schultz proposed an insider threat model that incorporated psychological profiling and behavioural observations such as disgruntlement, anger management and stress (Schultz, 2002). Other comprehensive insider-threat contributions have incorporated psychological traits such as OCEAN in addition to

behaviour and motivation (Legg et al., 2013) (Nurse et al., 2014). Such frameworks model insider threat by combining technological measures with human psychological traits to provide a more complete approach to the issue for better deterrence and detection.

Personality has been studied extensively through the years and was found related to many aspects of human work life such as job performance (Barrick and Mount, 1991). As it pertains to insider threats, personality traits such as Narcissism, Machiavellianism, and Excitement-Seeking were found related to insider threats and antisocial behaviour (Shaw et al., 2011) (Marcus and Schuler, 2004) (Axelrad et al., 2013). Moreover, OCEAN personality traits such as Openness could indicate that the individual is susceptible to phishing whilst Agreeableness and Excitement-Seeking traits have been correlated with antisocial behaviour (O'Connor and Dyce, 1998). In addition, researchers have attempted to predict personality of individuals in the online world. For example, Schwartz et al. studied 75,000 Facebook profiles and compared the language used in their profiles to identify distinctive words or phrases used by people based on demographics such as age, gender, location and psychological traits (Schwartz et al., 2013). Other contributions have defined a personality trait prediction model based on the textual analysis of the individual's Twitter profile (Golbeck et al., 2011a). Moreover, Yarkoni studied the personality and linguistic correlation in over 100,000 blogs and identified correlations between the personality traits and the LIWC dictionary (Yarkoni, 2010). Chorley et al. studied correlations between the OCEAN personality traits and location-based variables (e.g. number of places visited, venue popularity and so on) in location-based social networks (Chorley et al., 2015).

In terms of profiling the users Internet browsing interests, some contributions applied hierarchical clustering on websites an individual has visited to determine their long term and short-term interests (Grčar et al., 2005). Others aimed in determining the keywords that an individual may or may not be interested using a Term Frequency-Inverse Document Frequency (TF-IDF) and Neural Networks approach (Shaban et al., 2010). To the best of our knowledge, there has been no previous work on predicting an individual's personality traits based on the websites they visit. The efforts by Kosinski et al. have measured the personality of individuals and correlated them with website choice and found that there indeed exists a meaningful connection between an individual's personality and their website preference (Kosinski et al., 2014). However, that study was based on conducting personality tests, thus not automated. In this paper we extend on this idea by hypothesising that individuals' personality traits could be predicted by analysing the linguistic features of the websites they frequently visit and this could be used to determine the likelihood this individual might become an insider threat.

3 APPROACH

We hypothesise that similar to previous research that correlates an individual's personality traits with blog posts, Twitter feeds and Facebook profiles, we can infer an individual's personality based on the linguistic features of the websites they regularly visit. This could be used towards insider-threat detection by monitoring an individual's browsing pattern and creating their personality profile then assessing that profile for deviations in their personality traits over time. This profile deviation could then be incorporated in a more comprehensive insider-threat framework to determine whether the individual is a potential insider threat (Legg et al., 2013). However, we emphasis that monitoring employee's browsing activity could breach the employees privacy rights, therefore having legal and ethical implications (Kaupins and Minch, 2005). As such, this approach should be applied with care.

Our approach is composed of two main components. The Website–OCEAN personality correlation tool and the Insider-Threat application. We discuss both components in detail below.

3.1 Website–OCEAN Personality Correlation Tool

Here we explore the relationship between an individual's browsing interests and their OCEAN personality traits. We apply a category-based textual analysis approach using psychological groupings of linguistic features such as the LIWC (Linguistic Inquiry and Word Count) dictionary to identify links between website linguistic terms and personality traits (Pennebaker et al., 2001). LIWC is a widely used tool utilised in lexical approaches for personality measurement, and produces statistics on 81 different categories (such as Social, Work, Money and so on) by calculating the percentage of words in the input that match pre-defined words in a given category. The LIWC dictionary categories represent the features in our approach, most of which contain hundreds of words. We apply k-means clustering to identify what websites are associated with each OCEAN trait.

Figure 1 shows the workflow that we shall adopt, as described below.

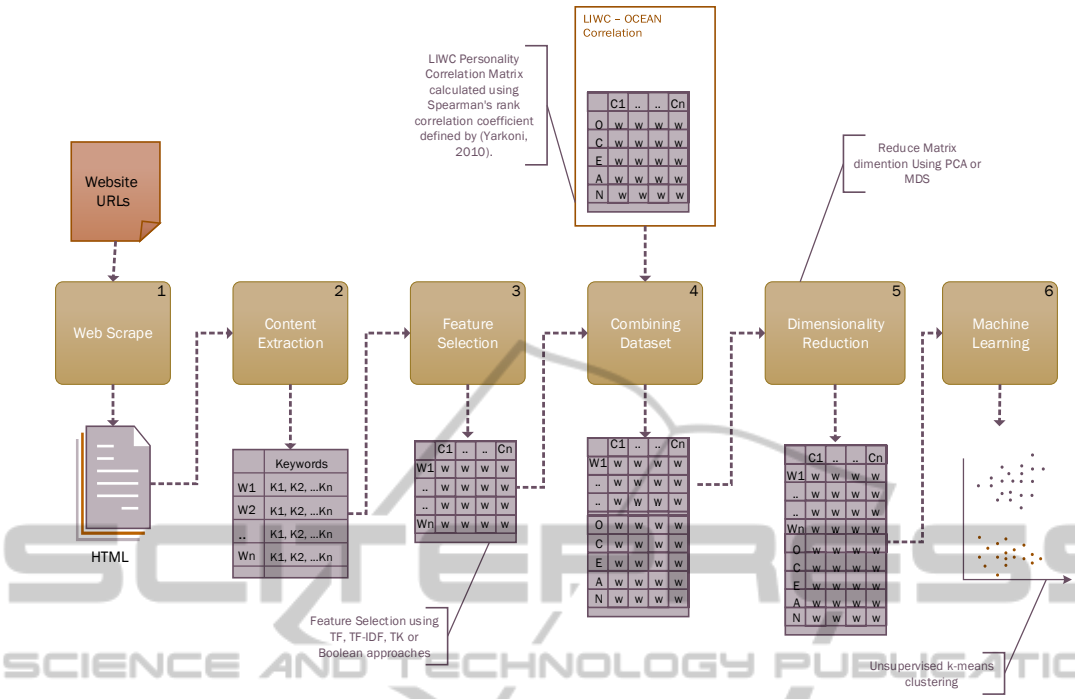

Figure 1: Website–OCEAN Personality Correlation tool process: (1) Website is scraped to extract textual content; (2) The website words are processed to remove any HTML characters, websites are represented as vector of words; (3) Website words are mapped to the LIWC features (categories) and features with the highest weight are selected; (4) The website–LIWC matrix and OCEAN–LIWC matrix is combined; (5) Reduce dimensions of resulting matrix using PCA or MDS; and (6) Apply an unsupervised machine learning method (*k*-means) on resulting matrix to identify Website–OCEAN correlations.

1. **Web Scraping.** Websites are a combination of textual and contextual features such as images, hyperlinks, and HTML tags and scripts. We apply the text-only website classification approach in our analysis, where we only consider the textual content for the website's home page (e.g, `index.html`).

2. **Content Extraction and Pre-processing.** Firstly, we extract the textual content from the website. In this particular context, *content* refers to the words in the HTML document for that particular website and include the website title, the element enclosed in the `<title></title>` HTML tags, website keywords enclosed in the HTML element where `id=Keywords`, in addition to the body content text. During website extraction phase, we flag each extracted word if it is a term in the title or is a keyword. Each website is represented as word-vectors (also called a 'Bags of words' approach). For example, terms and thus the word-vector for a money loaning website may include `Money`, `loan`, `debt` and so on. Once the website terms are determined, the terms are pre-processed to remove any JavaScript and HTML tags. Stop words are removed from the word-vector for example `a`, `the`, `and`, `to` and so on. Moreover, technical terms used in all websites regardless of the website class such as `form`, `sign-in`, `contact`, `basket`, `home` were removed as they would add little value to the analysis and are frequently used in websites. In addition, words that are less than three characters are removed. Although Yarkoni indicated that many words with the same stem had quite different patterns of correlation with personality (e.g., `love` and `lover`), most of the LIWC categories include all words that share a particular stem, and thus we preform stemming in our analysis (Yarkoni, 2010).

3. **Feature Selection.** Feature selection is the process of selecting a subset of *good* features, that removes noise whilst preserving informative features for our task. We first construct a matrix representation of the textual content that maps the websites terms to the LIWC categories (features). For each website, we compute the frequency of each term that matches LIWC predefined dictionary categories. For example, a bank website might include terms such as *hire* or *tax* that maps to the LIWC *Work* category as well

as *debt* or *loan* that map to the *Money* category. In our approach, we compute a weight score for each website–LIWC category mapping to capture the relevancy among websites and LIWC linguistic categories. The highest weighted LIWC categories from the feature space are selected. We define our features using four selection approaches described below.

- Term Frequency approach (TF): This approach is simply counting the number of words that belong to each LIWC category for each website.
- Term (word) Frequency/ Inverse Document Frequency (TF-IDF) approach:

Each website–LIWC mapping is assigned a weight based on the product of Term Frequency (TF) the number of times a word W occurs in the page and Inverse Document Frequency (IDF). TF-IDF is defined in the following equation:

$$W = \frac{FWF}{max(FWF)} \times log\frac{NW}{NWW} \quad (1)$$

where FWF is the frequency of word in the website, $max(FWF)$ is the maximum frequency of website containing the word, NW is the total number of websites in dataset, and NWW is the number of websites containing a word in the LIWC category.

The resulting weight is proportional to the frequency of the category words within the website, thus a category that contain words that appear frequently on a website is considered to be more important. Moreover, by dividing the frequency of the category words by the maximum frequency of word on same page, we normalize the word use that could be meaningfully compared across websites. Therefore, words that are common in all websites will be weighted low, which is advantageous at filtering out any propositions, articles and pronouns (Shaban et al., 2010).

- Title-Keywords (TK) weight approach: The topic of a website can often be determined by examining the website title or keywords. Thus, words that appear in the website title or keyword list of the HTML structure will be assigned a higher weight (e.g., multiplied by 3). In our approach, we applied the following weighting:

$$W = FWF(3t + 3k) \quad (2)$$

where FWF is the frequency of the word in the website, $k = 1$ if word is a keyword and 0 otherwise, and $t = 1$ if word is in the website title and 0 otherwise.

- Boolean approach: If the frequency of words in a LIWC category more than 5, then we assign it a weight 1 otherwise the weight is 0.

In addition, we excluded some LIWC categories in which website terms did not fall within to eliminate what is likely be noise (such as pronouns, verbs present-tense and future-tense, and negations). Moreover, we removed website observations where the sum of terms mapped to the LIWC categories were less than five, this is to eliminate what could be noise in our dataset.

4. **Combining Dataset.** The websites–LIWC category feature set will be represented as a matrix of category weights. This is combined with the LIWC–OCEAN correlation as defined by Yarkoni (Yarkoni, 2010). The dataset is then normalized to values between ± 1.
5. **Dimensionality Reduction.** The result matrix combines the websites and psychological traits as a matrix of normalized category weights. This is represented by a large number of features that result in a sparse matrix. Some of these features may be considered irrelevant or redundant, and thus can be removed from the matrix. Dimensionality reduction aims to reduce the feature space using statistical means that result in a lower dimension matrix. In our approach, we apply the dimensionality reduction techniques Principle Component Analysis (PCA) (Jolliffe, 2005) and Multidimensional Scaling (MDS) (Kruskal and Wish, 1978).
6. **Machine Learning.** The matrix is then input into unsupervised machine learning method to train a model and define the individual's browsing behaviour and personality correlations. Specifically, k-means is applied to cluster the websites and the OCEAN personality traits. The clustering will determine the personality traits associated to each website.

3.2 Insider-threat Application

In the aim of preventing insider threats, organisations could utilise technological means that monitor their employees Internet browsing activities and record the websites they visit. Using our approach as illustrated in Figure 2, each employee will have a browsing profile that includes what websites the individual accessed, timestamp, number of times website has been accessed and time spent on the page. This browsing profile is then input to the Website and Personality Correlation tool defined in Section 3.1. For each website, the Euclidean distance between the website cen-

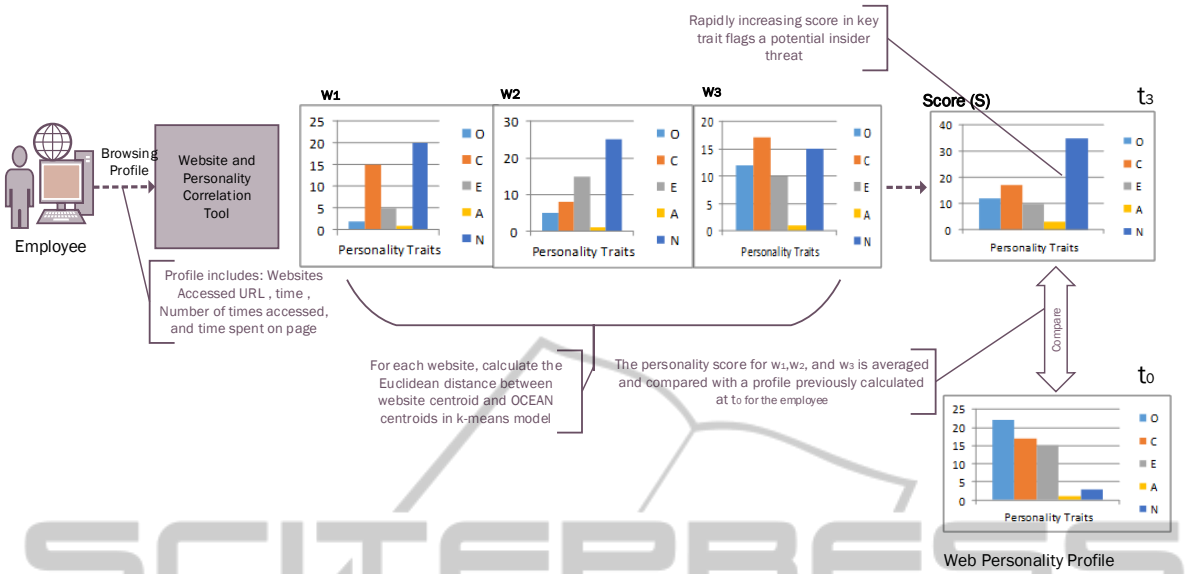

Figure 2: Insider-Threat Detection Application: Employees browsing profile is input into our tool. The tool then calculates the OCEAN personality trait scores for each website based on the websites' textual content. The score for each OCEAN personality trait is averaged. By comparing each averaged OCEAN score with the previously calculated score in the employees personality profile, we can detect personality deviations. In this example, the Neuroticism score has increased rapidly from 3 to 35, which indicates a significant change in behaviour and possibly a potential insider threat.

troid and OCEAN personality trait centroid is measured using the *k*-means model. The personality score *S* for each OCEAN personality trait is then calculated using the following equation:

$$S = \frac{w_1c_1 + w_2c_2 + \ldots\ldots + w_nc_n}{n} \quad (3)$$

where w is the OCEAN personality trait score for a website, c is the number of times the website was accessed or the time spent on page, and n is the number of websites in the browsing profile.

The resulting personality score for each trait is then compared to the employee's web personality profile that contains previously calculated personality scores. This comparison is done to determine how well the new websites observations fit with the currently defined profile for that particular individual. As a result, we can detect any pattern deviations or rapidly changing personality scores. We hypothesise that there exists a strong correlation in personality traits and the likelihood of becoming an insider threat. In addition, deviations in personality traits resulting from changes in sites visited is a potentially viable method at detecting insider threats. For example, significant changes in personality traits has an impact on the likelihood of becoming insider threat.

The personality profile can then be used as a metric in a comprehensive insider-threat framework and aid in detecting employee browsing behaviour deviations that can potentially be an insider-threat risk. For example, this can be incorporated in the `Mindset` element in the `People Lane` of the model defined in our previous work (Legg et al., 2013).

4 EXPERIMENT

4.1 Dataset

The website dataset was generated by extracting the contextual features of 1000 randomly selected website samples categorised to the 15 different categories defined by The Open Directory Project (DMOZ) (DMOZ, nd). These categories are Adult, Arts, Business, Computers, Games, Health, Home, Kids, News, Recreation, Reference, Science, Shopping, Society, and Sports. The DMOZ 'World' category was not considered as it contains websites in non-English languages. The personality dataset used was based on the findings of Yarkoni, where OCEAN personality traits were correlated to LIWC categories using Spearman's rank correlation coefficient (Yarkoni, 2010). The OCEAN-LIWC correlation values were normalized to values between ± 1. Each of the OCEAN domains has a cluster of correlated and more specific lower facets (personality traits) (Goldberg, 1999). For instance, Extroversion includes such related qualities as Friendliness, Gregariousness, Assertiveness, Excitement-seeking, Activity-level, and

Cheerfulness. We consider the OCEAN main domains in addition to their more specific personality traits in our approach.

4.2 Experiment Setup

In order to evaluate the proposed approach, we conduct a number of experiments using the data of 1000 websites that are based on the different combinations of feature selection and dimensionality reduction techniques. This allows us to compare the different feature selection approaches in terms of relevance to the task of website and personality correlation, and also to derive a greater understanding of website interests of different personality types. We employ unsupervised machine learning techniques to be able to build an individual's Internet browsing model that maps website accessed to personality traits. The proposed system was built using Python and the KNIME data analysis software (Surhone, 2010). Python was primarily used for scraping the website content, determining content terms and pre-processing of the dataset. The resulting website–LIWC categories matrix was then imported into KNIME along with the normalized OCEAN–LIWC correlation matrix to apply machine learning techniques and evaluate the results. For each dataset, the following experiments were conducted:

1. Website–OCEAN clustering: We applied unsupervised machine learning methods i.e., k-means clustering (where $k = 5$) to establish similar website and personality traits groupings on the complete website and personality dataset. We then computed the distance matrix, which is the euclidean distance between each website centroid and OCEAN personality trait centroid.
2. Website–OCEAN Dot Product: We computed the dot product matrix of the website–LIWC matrix and the OCEAN–LIWC correlation matrix to determine the degree on similarity between these matrices. (Yarkoni, 2010).

5 RESULTS

There are a number of key results from our experiment. In terms of the dimensionality reduction methods, both PCA and MDS had similar reduction results although the PCA execution time was slightly faster in KNIME. Our k-means clustering experiment has shown that the boolean feature selection approach achieved the most suitable clustering of websites and personality traits. Other feature selection approaches resulted in websites centroids and personality centroids to separate from each other, thus personality traits where all assigned to one cluster whilst websites were assigned to other clusters. This may be due to the difference of how the website–LIWC category matrix was calculated in the non-boolean feature selection approach compared to the OCEAN–LIWC correlation matrix. An attempt was made to normalize these values, however this did not improve the clustering results. We show the cluster of websites and personality traits using the boolean feature selection approach in Figure 3. To gain an understanding on how each website relates to the personality traits, we also calculated the euclidean distance matrix of the k-means clustering between the interesting website observation and each personality trait.

The lower facets for each of the OCEAN main domains have clustered correctly around the main OCEAN trait. For example, the lower facets of Neuroticism, such as Depression, Anxiety, Self-consciousness, Hostility, Vulnerability are all in the Neuroticism cluster. Similarly, the lower facets for Openness and Conscientiousness all cluster around their main OCEAN trait. Extraversion and Agreeableness are both in one cluster as well as their lower facets although some of their lower facets were found in another cluster. We do not have a reason as to why this occurred but by observing the dot product results, we can see that Extraversion and Agreeableness have similar values and thus the cause may be the values in the LIWC–OCEAN correlation matrix.

The dot product similarity measure results for a sample of website observations and personality traits observations is shown in Figure 4 as a heat map. By using the dot product measure, we are able to compare each OCEAN personality trait with each website. The results indicate a reasonable correlation between website content and OCEAN personality traits. For example, websites categorised by DMOZ as `News/Analysis_and_opinion` websites such as `themedialine.org` and `projectsyndicate.org` scored high in the OCEAN domain Neuroticism and its constituent specific personality traits (Hostility, Depression, Self-Consciousness and Anxiety) in addition to Extraversion's specific trait (Excitement-seeking). They also scored low in Sympathy, Imagination, and Cautiousness. Weight-loss websites scored high in Anxiety, Depression and Altruism whilst scored low in Assertiveness and Intellect. Moreover, `controllingparents.com`, a website dedicated to helping support individuals who have controlling or narcissistic parents, scored high in Neuroticism and its more detailed facets. Sports websites scored high in Conscientiousness and Open-

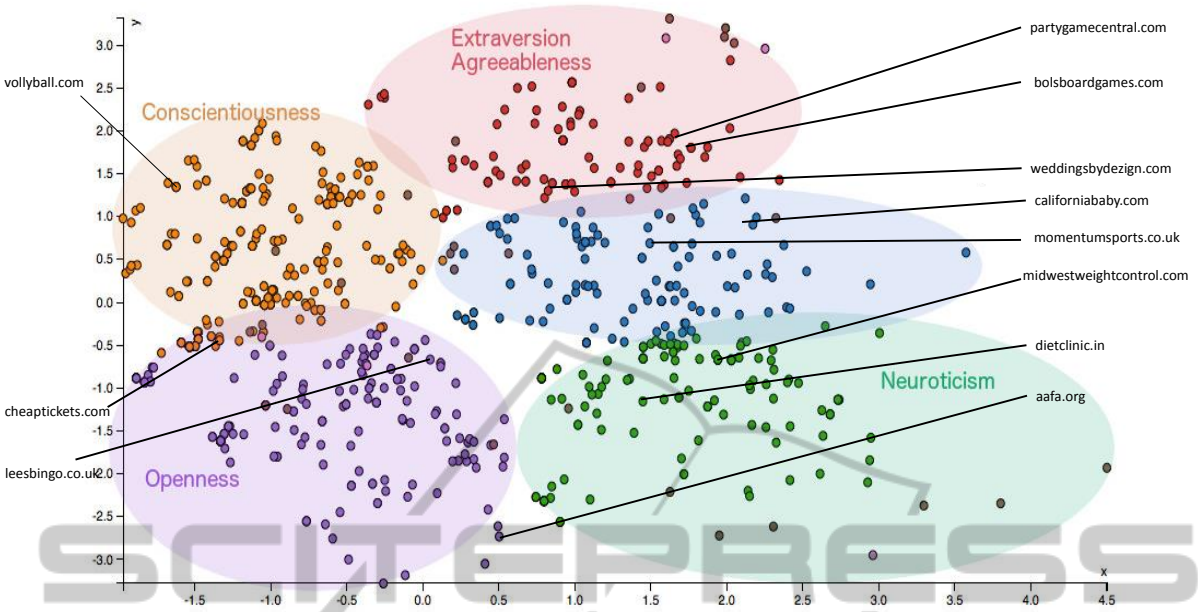

Figure 3: Website and Personality Clustering using *k*-means and the boolean feature selection approach on the PCA reduced dataset.

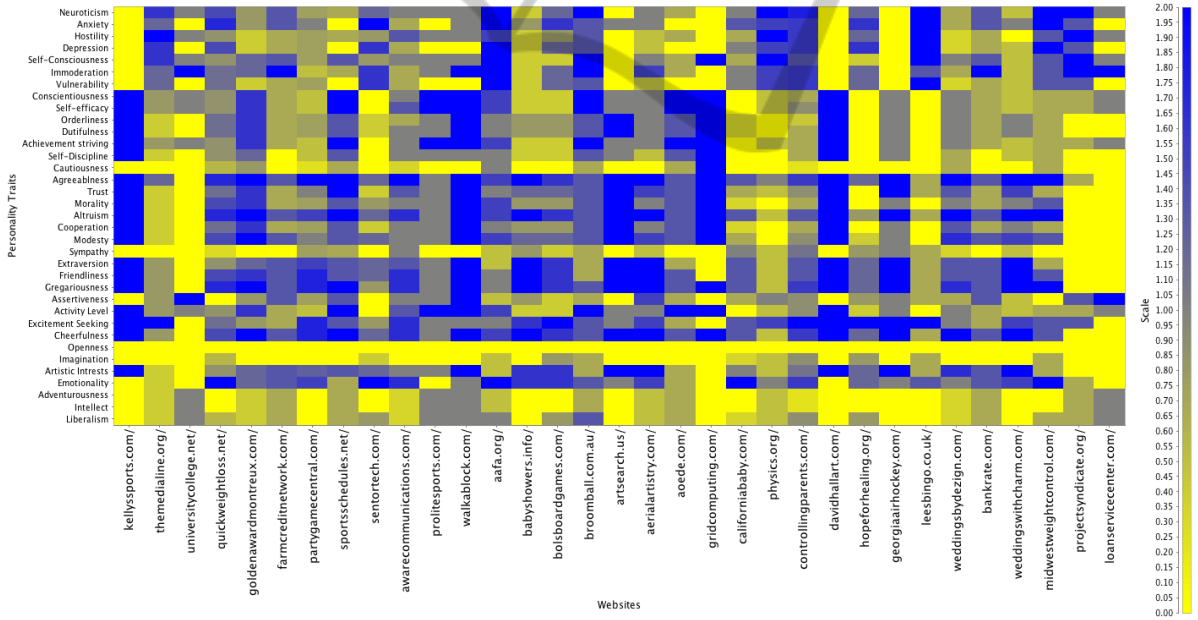

Figure 4: Dot product similarity measure of the website–LIWC matrix and OCEAN–LIWC Spearsman's rank correlation as defined in (Yarkoni, 2010). Darker shades (blue) indicate a positive correlation while lighter shades (yellow) indicate a negative correlation. For example, `leesbingo.co.uk`, a gambling website, is positively correlated with Neuroticism and its more specific traits (Anxiety, Hostility, Depression, Self-consciousness, Immoderation and Vulnerability) in addition to Excitement-seeking whilst negatively correlated with Adventurous, Intellect and Liberalism.

ness and other facets such as Assertiveness, Activity-level, Self-discipline, Achievement-striving, Orderliness and Dutifulness. They also scored low in Depression, Anxiety, Vulnerability and Imagination. Wedding websites were found to be highly associated with Agreeableness and Openness and lower facets such as Emotionality and Cheerfulness. Money loan websites such as `loanservicecenter.com`, `bankrate.com`, and `walkablock.com` scored high in Immoderation and Assertiveness. Finally,

artistic websites such as davidhallart.com and artsearch.us scored high in Artistic-interest. We summarise some of the findings in Table 1 and show the keywords and LIWC category mappings that led to these personality scores.

6 DISCUSSION

6.1 Results

In this paper we hypothesised that an individual's browsing interests can lead to inferences about their personality traits. By monitoring these traits as they are, and also as they change over time, we believe that they could be a novel way to identify potential insider threat. We conducted various experiments using different variations of feature selection and dimensionality reduction methods. The result of our experiments provide meaningful knowledge and understanding. For example, we can imagine if a person is athletic and is interested in sport websites they might be Energetic, Adventurous, Sociable, Persistent and Self-disciplined thus, score high on Extraversion and Consciousness whilst scoring low in Neuroticism. These results are similar to the findings of Hamburger and Ben-Artzi (Hamburger and Ben-Artzi, 2000). Moreover, people who are unhappy with their weight and are frequently searching online for solutions can be anxious and maybe depressed, thus scoring high in Neuroticism similar to the findings of other contributions (Davis and Fox, 1993).

We also have reached similar findings as to the study by Yarkoni (Yarkoni, 2010). For example, hockey websites such as georgiaairhockey.com, broomball.com and physics websites such as physics.org scored high in Neuroticism. Business-marketing websites such as goldenawardmontreus.com and awarecommunications.com scored low in Neuroticism. Health-disease websites such as aafa.org scored high in Agreeableness. In addition, artistic websites such as davidhallart.com and aerialartistry.com score high in Artistic-interest similar to the findings by Yarkoni (Yarkoni, 2010), however they scored low in Liberalism. Although previous research support our results, a thorough evaluation that would involve a large user study will be conducted in the future.

In terms of the methodology used, correlation between personality traits and LIWC categories were based on the findings of Yarkoni that correlated LIWC categories to word use in blogs (Yarkoni, 2010). A more effective approach would be to combine website textual features and LIWC correlations with psychological interview scores and investigate the correlation between websites and personality. Then, correlation findings could be used in the defined model to gain more accurate results.

In addition, the LIWC dictionary was initially designed to predict the personality traits of individuals based on their writings, such as personal blog posts, Twitter, Facebook and written articles and essays. We applied LIWC to identify the personality of people that *view* the website textual content rather the personality of the person that wrote them. As a result, a number of LIWC categories were removed from the analysis such as the use of pronouns and articles which are important categories in analysing written text rather than viewed text. This has affected some of the correlation values in the correlation matrix values adopted. For example, Openness was highly correlated with LIWC categories that were removed from our analysis, and negatively related with the other LIWC categories. Therefore, as shown in the dot product heat map (Figure 4), the Openness value in all correlation was equal to zero. Therefore, defining a new dictionary that maps personality traits into observed context rather than written could prove a useful step for future work. We also applied the category-based approach in text analysis that is based on mapping each website word into a defined category, such as the LIWC categories. However this approach may limit the findings to preconceived relationships with words or categories as many of the websites words did not fall into any of the LIWC categories, thus were not considered in the analysis. An open vocabulary approach, such as the one defined by Schwartz et al. may be useful to experiment with in the future (Schwartz et al., 2013).

Furthermore, since research relies heavily on text analysis, the nature of text on websites raises an interesting challenge. Modern websites are becoming more dynamic thus their content may change everyday. Other challenges include filtering advertisements and irrelevant terms that interfere with the accuracy of the results. In addition to noisy input data, the usage of an unsupervised clustering may limit the intelligence of the system. The core motivation for applying k-means clustering is that there is no existing ground truth on what websites are associated with which personality traits. Although the research by Kosinski et al. did provide some insight on what OCEAN personality traits are associated with websites, the websites list provided was not sufficient and the textual content scraped of the websites were not enough, thus filtered out in the pre-processing phase of our approach (Kosinski et al., 2014). In addition, personality classification is a multi-class problem, therefore adopt-

Table 1: Website keywords and the associated LIWC Categories mappings based on the website keywords.

DMOZ Category	Websites	Keywords	LIWC Categories
Health\Weight-Loss	quickweightloss.net dietclinic.in midwestweightcontrol.com	e.g. loss, control, goal, diet	Sadness Achievement, Affective Processes Negative Emotions
Money\Finance	loanservicecenter.com farmcreditnetwork.com bankrate.com	e.g. loan, insurance, bank, grant, free	Money Work Affective Processes
Arts	davidhallart.com artsearch.us aerialartistry.com	e.g gallery, artist, fair, planner	Leisure Positive Emotions
News\Analysis_and_Opinion	projectsyndicate.org themedialine.org	e.g. news, newspaper, politics, affairs	Space Cognitive Processes Insight Work
Society\Support_Groups	drirene.com hopeforhealing.org controllingparents.com	e.g. emotional, counseling, abuse, violence, addiction	Anger Affective Processes Negative Emotions
Games\Party_Games	bolsboardgames.com gatheraroundgames.com partygamesetc.com	e.g. social, fun, party, family, friends	Social Processes Leisure Affective Processes Positive Emotions
Shopping\Weddings Home\Entertaining	weddingsbydezign.com weddingswithcharm.com babyshowers.info	e.g. baby, favors, planning, gift, party, presents, games, flowers, wedding	Social Processes Leisure Affective Processes Positive Emotions
Sports	kellyssports.com sportsschedules.net vanguardathletics.com prolitesports.com	e.g. baseball, football, running, exercise	Leisure
Sports\Hockey	broomball.com.au leaguelineup.com georgiaairhockey.com	e.g. match, goal, leagues, tournaments	Social Processes Leisure
Science\Physics	physics.org	e.g. school, university, high, question, top	Social Processes Cognitive Processes Insight Work

ing a multi-class classification approach may be better suited for this task.

6.2 Insider-threat Application

The important question that comes from this research is how the results can be used. Drawing on previous research, there is potential to integrate personality-trait prediction results into a comprehensive insider-threat framework to enhance the accuracy of insider-threat risk prediction. For example, personality traits such as Narcissism, Machiavellianism, and Excitement Seeking have been found to be related to insider threats and antisocial behaviour (Shaw et al., 2011) (Marcus and Schuler, 2004) (Axelrad et al., 2013). In addition, individuals found to be narcissistic display an excessive self-importance, strong need for admiration and lack of empathy—recurring traits in individuals who committed insider attacks (Shaw et al., 1998). Moreover, OCEAN personality traits such as Openness can indicate that the individual could be susceptible to phishing spam. Research also found that people who commit insider attacks tend to be introverted individuals, a trait that was found highly presented in IT specialist (Pocius, 1991). Our results could support such knowledge, since technology based websites `sentortech.com` and especially `gridcomputing.com` (that target tech savvy people) do indeed score low in Extraversion.

The personality prediction results can be used to build an individual's browsing profile that includes their website interests and predicted personality traits over a period of time. Using such a profile, an insider-threat framework could monitor the individual's browsing and personality behaviour over time and detect any deviation in their normal pattern. Although personality traits are useful in predicting insider-threat risk potential, they are not sufficient in isolation to identify a potential attacker (Pocius, 1991), since other factors such as environment and

precipitating events need to be considered also. For example, an individual who scores high in Neuroticism might be the perfect candidate for a job position; however, when combined with other behavioural factors (e.g. stress) or the individual was offered a better position in another organisation than could increase the likelihood of threat. Moreover, cluster of personality characteristics might be useful to consider (Shaw et al., 1998). However, we note that monitoring employees' Internet activities may be deemed as a breach of privacy and therefore may have legal and ethical implications that organisations need to consider.

7 CONCLUSIONS AND FUTURE WORK

In this paper we explored the possibility of predicting an individual's personality traits based on their browsing history in order to detect potential insider threats. We conducted various experiments from *k*-means clustering to dot product similarity measure, in order to identify possible relationships between websites and OCEAN personality traits. The results were well matched and supported by previous research. We believe that our approach could be integrated in a comprehensive insider-threat framework by observing an individual's browsing deviations over time and monitoring for personality changes that may allude to the individual being a potential insider threat.

The next step will be to evaluate the results by conducting a study where individuals are assessed to determine their actual personality trait scores and that is compared with the output of our approach based on their browsing behaviour. Furthermore, the content of websites accessed were only considered in our analysis. We believe that we can achieve more accurate profiling results on what websites an individual is interested in by considering other factors such as the time spent on the page and the number of times the website was accessed (Liang and Lai, 2002).

ACKNOWLEDGEMENTS

We are grateful to the Ministry of Higher Education in the Kingdom of Saudi Arabia, the Saudi Arabian Cultural Bureau in London, and King Saud University for their financial sponsorship and support of Bushra Alahmadi's DPhil programme.

REFERENCES

Allport, G. W. (1962). The general and the unique in psychological science1. *Journal of personality*, 30(3):405–422.

Axelrad, E. T., Sticha, P. J., Brdiczka, O., and Shen, J. (2013). A bayesian network model for predicting insider threats. In *Security and Privacy Workshops (SPW), 2013 IEEE*, pages 82–89. IEEE.

Barrick, M. R. and Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1):1–26.

Cappelli, D. M., Moore, A. P., and Trzeciak, R. F. (2012). *The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes*. Addison-Wesley.

Chorley, M. J., Whitaker, R. M., and Allen, S. M. (2015). Personality and location-based social networks. *Computers in Human Behavior*, 46(0):45 – 56.

Davis, C. and Fox, J. (1993). Excessive exercise and weight preoccupation in women. *Addictive Behaviors*, 18(2):201 – 211.

DMOZ (n.d). Open Directory Project, http://www.dmoz.org.

Ehrman, K., Jagid, B., and Loosmore, N. B. (1997). Electronic control system/network. US Patent 5,682,142.

Golbeck, J., Robles, C., Edmondson, M., and Turner, K. (2011a). Predicting personality from Twitter. In *Privacy, security, risk and trust (PASSAT), 2011 IEEE third international conference on and 2011 IEEE third international conference on social computing (socialcom)*, pages 149–156. IEEE.

Golbeck, J., Robles, C., and Turner, K. (2011b). Predicting personality with social media. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 253–262. ACM.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*, 7:7–28.

Grčar, M., Mladenić, D., and Grobelnik, M. (2005). User profiling for interest-focused browsing history. In *In SIKDD 2005 at Multiconference IS 2005*.

Greitzer, F. L. and Frincke, D. A. (2010). Combining traditional cyber security audit data with psychosocial data: towards predictive modeling for insider threat mitigation. In *Insider Threats in Cyber Security*, pages 85–113. Springer.

Hamburger, Y. and Ben-Artzi, E. (2000). The relationship between extraversion and neuroticism and the different uses of the internet. *Computers in Human Behavior*, 16(4):441 – 449.

Hunker, J. and Probst, C. W. (2011). Insiders and insider threats—an overview of definitions and mitigation techniques. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 2(1):4–27.

Jolliffe, I. (2005). *Principal component analysis*. Wiley Online Library.

Kaupins, G. and Minch, R. (2005). Legal and ethical implications of employee location monitoring. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*, pages 133a–133a. IEEE.

Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., and Graepel, T. (2014). Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning*, 95(3):357–380.

Kruskal, J. B. and Wish, M. (1978). *Multidimensional scaling*, volume 11. Sage.

Legg, P., Moffat, N., Nurse, J. R. C., Happa, J., Agrafiotis, I., Goldsmith, M., and Creese, S. (2013). Towards a conceptual model and reasoning structure for insider threat detection. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 4(4):20–37.

Liang, T.-P. and Lai, H.-J. (2002). Discovering user interests from web browsing behavior: An application to internet news services. In *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on*, pages 2718–2727. IEEE.

Marcus, B. and Schuler, H. (2004). Antecedents of counterproductive behavior at work: a general perspective. *Journal of Applied Psychology*, 89(4):647.

Nord, G. D., McCubbins, T. F., and Nord, J. H. (2006). E-monitoring in the workplace: privacy, legislation, and surveillance software. *Communications of the ACM*, 49(8):72–77.

Nurse, J. R. C., Buckley, O., Legg, P. A., Goldsmith, M., Creese, S., Wright, G. R., and Whitty, M. (2014). Understanding insider threat: A framework for characterising attacks. In *Workshop on Research for Insider Threat (WRIT) held as part of the IEEE Computer Society Security and Privacy Workshops (SPW14), in conjunction with the IEEE Symposium on Security and Privacy (SP)*. IEEE.

O'Connor, B. P. and Dyce, J. A. (1998). A test of models of personality disorder configuration. *Journal of Abnormal Psychology*, 107(1):3.

Paulhus, D. L. and Williams, K. M. (2002). The dark triad of personality: Narcissism, machiavellianism, and psychopathy. *Journal of research in personality*, 36(6):556–563.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71:2001.

Phyo, A. and Furnell, S. (2004). A detection-oriented classification of insider IT misuse. In *Third Security Conference*.

Pocius, K. E. (1991). Personality factors in human-computer interaction: A review of the literature. *Computers in Human Behavior*, 7(3):103–135.

PWC (2014). US cybercrime: Rising risks, reduced readiness: Key findings from the 2014 US state of cybercrime survey.

Schultz, E. E. (2002). A framework for understanding and predicting insider attacks. *Computers & Security*, 21(6):526–531.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E., et al. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791.

Shaban, K. B., Chan, J., and Szeto, R. (2010). Interest-determining web browser. In *Advances in Data Mining. Applications and Theoretical Aspects*, pages 518–528. Springer.

Shaw, E., Ruby, K., and Post, J. (1998). The insider threat to information systems: The psychology of the dangerous insider. *Security Awareness Bulletin*, 2(98):1–10.

Shaw, E. D., Stock, H. V., et al. (2011). Behavioral risk indicators of malicious insider theft of intellectual property: Misreading the writing on the wall. *White Paper, Symantec, Mountain View, CA*.

Shen, J., Brdiczka, O., and Liu, J. (2013). Understanding email writers: Personality prediction from email messages. In *User Modeling, Adaptation, and Personalization*, pages 318–330. Springer.

Spitzner, L. (2003). Honeypots: Catching the insider threat. In *Computer Security Applications Conference, 2003. Proceedings. 19th Annual*, pages 170–179. IEEE.

Sumner, C., Byers, A., Boochever, R., and Park, G. J. (2012). Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, volume 2, pages 386–393. IEEE.

Surhone, L. M. (2010). *KNIME: R (programming Language), WEKA, Java*. Betascript Publishing.

Urbaczewski, A. and Jessup, L. M. (2002). Does electronic monitoring of employee internet usage work? *Communications of the ACM*, 45(1):80–83.

Wiggins, J. S. (1996). *The five-factor model of personality: Theoretical perspectives*. Guilford Press.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, 44(3):363–373.